Jumistik
E-ISSN: 2964-3953 P-ISSN: 2964-3888 Jurnal Manajemen Informatika, Sistem Informasi dan Teknologi Komputer

Vol.4 No.1 Juni 2025
Hal 430-441
Https://ojs.amiklps.ac.id

ANALISIS SENTIMEN TERHADAP TERORISME PADA *PLATFORM TWITTER* MENGGUNAKAN *SUPPORT VECTOR MACHINE*

Novriza Rahayu[1*], Sylvia Indri Yani[2], Marwah[3], Andre Pratama[4]
[1,2,3,4] Informatika, Universitas Satya Terra Bhinneka
E-mail: novrizarahayu18@gmail.com[1*], sylviaindriyani1321@gmail.com[2], marwah3085@gmail.com[3], andrepratama@satyaterrabhinneka.ac.id[4]

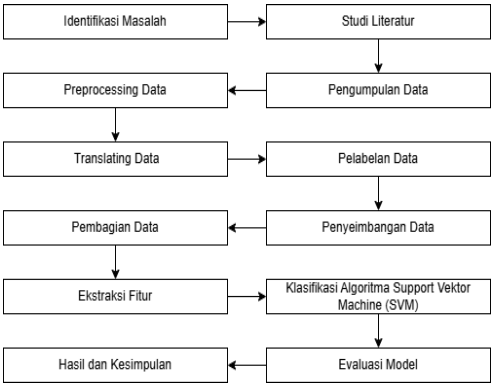
INFO ARTIKEL

Sejarah Artikel
Diterima : 23/06/2025
Direvisi : 28/06/2025
Diterbitkan : 29/06/2025

***Corresponding author**
novrizarahayu18@gmail.com

DOI: 10.70247/jumistik.v4i1.152
Https://ojs.amiklps.ac.id

GRAPHICAL ABSTRACT



ABSTRACT

This study aims to classify public opinions on terrorism issues using the Support Vector Machine (SVM) algorithm. The initial data, consisting of texts in Indonesian, underwent several stages, including preprocessing, translation into English, and labeling using VADER. Data imbalance was addressed using Random Over Sampling, while numerical representation was obtained through Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction. The model was evaluated using a confusion matrix and k-fold cross-validation to ensure its performance and generalization. The confusion matrix results showed an accuracy of 98.02%, precision of 98.09%, recall of 98.02%, and an f1-score of 98.01%. Meanwhile, the k-fold cross-validation with k = 5 demonstrated an average accuracy of 97.77%, precision of 97.86%, recall of 97.77%, and an f1-score of 97.76%. This approach significantly outperformed the Long Short-Term Memory (LSTM) method, which achieved only 59.34% accuracy on similar issues. The combination of preprocessing, data balancing, feature extraction, and validation processes proved effective in creating an accurate sentiment classification model, contributing significantly to the development of text-based sentiment analysis technology to support decision making.
Keywords: *Sentiment Analysis, Support Vector Machine, Terrorism, Twitter*

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan pandangan masyarakat tentang isu terorisme menggunakan algoritma *Support Vector Machine* (SVM). Data awal yang berupa teks dalam bahasa Indonesia akan melalui beberapa tahap, yakni *preprocessing*, penerjemahan ke dalam bahasa *Inggris*, serta pelabelan menggunakan VADER. Ketidakseimbangan data ditangani dengan menggunakan *Random Over Sampling*, sementara representasi numeriknya diperoleh melalui ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF). Model dievaluasi menggunakan *confusion matrix dan k-fold cross-validation* untuk memastikan performa dan generalisasi. Hasil *confusion matrix* menunjukkan *accuracy* 98,02%, *precision* 98,09%, *recall* 98,02%, dan *f1-score* 98,01%. Evaluasi *k-fold cross-validation* dengan k = 5 menunjukkan rata-rata *accuracy* sebesar 97,77%, *precision* 97,86%, *recall* 97,77%, dan *f1-score* 97,76%. Pendekatan ini menunjukkan keunggulan signifikan dibandingkan metode *Long Short-Term Memory* (LSTM) yang hanya memperoleh *accuracy* 59,34% pada isu yang sama. Kombinasi dari metode *preprocessing*, penyeimbangan data, ekstraksi fitur, dan proses validasi terbukti efektif dalam menciptakan model klasifikasi sentimen yang akurat, yang memberikan kontribusi penting untuk pengembangan teknologi analisis sentimen berbasis teks dalam mendukung pengambilan keputusan.
Kata kunci: Analisis Sentimen, *Support Vector Machine*, Terorisme, Twitter

PENDAHULUAN

Media sosial telah berkembang menjadi platform digital yang menekankan eksistensi dan partisipasi aktif penggunanya. Melalui media sosial, individu dapat membentuk identitas diri, berinteraksi, berkolaborasi, serta menjalin hubungan sosial secara virtual. Peranannya sebagai ruang interaktif di internet menjadikannya medium penting dalam membangun komunikasi dan jejaring di era modern [1]. Twitter termasuk jejaring sosial populer yang kerap dimanfaatkan publik untuk mengungkapkan pandangan mereka tentang beragam isu, termasuk politik, ekonomi, kasus korupsi, hingga isu sensitif seperti terorisme. Dengan fitur-fitur seperti *trending topics*, balasan (*replies*), dan *retweet*, platform ini menawarkan sumber data yang luas, dinamis, dan relevan untuk menganalisis opini publik secara mendalam dan terarah [2].

Terorisme merupakan tindak kejahatan yang sangat serius dan menjadi fokus perhatian global, tak terkecuali di Indonesia. Aksi-aksi teror yang terjadi belakangan ini di Indonesia tidak hanya berdimensi kriminal, tetapi juga memiliki keterkaitan dengan aspek ideologi, sejarah, dan politik, serta mencerminkan dinamika lingkungan strategis di tingkat regional maupun global. [3]. Dalam konteks ini, isu terorisme di Indonesia menarik perhatian dunia internasional karena dianggap menyimpang dari ekspektasi umum. Meskipun Indonesia menduduki posisi negara dengan jumlah penduduk Muslim terbesar di dunia, masyarakatnya justru menunjukkan sikap tegas dalam mengecam dan menolak segala bentuk aksi terorisme, sekaligus mendukung upaya kolektif untuk menciptakan perdamaian dan harmoni sosial [4].

Media sosial seperti Twitter menghasilkan volume data yang sangat besar dan beragam, yang mencerminkan persepsi, emosi, dan tanggapan masyarakat terhadap peristiwa atau tokoh tertentu. Namun, data yang diperoleh dari platform ini umumnya tidak terstruktur, bersifat singkat, serta mengandung elemen khas media sosial seperti emotikon, singkatan, sarkasme, dan bahasa informal. Karakteristik tersebut menuntut penerapan metode analisis yang lebih canggih dan sistematis untuk memperoleh pemahaman yang akurat [5]. Selain itu, dalam konteks terorisme, diskusi di media sosial seringkali diwarnai dengan sentimen yang sangat polarisasi, mulai dari kemarahan, ketakutan, empati terhadap korban, hingga perdebatan politik mengenai kebijakan keamanan dan kontraterorisme. Kompleksitas ini memerlukan pendekatan analisis yang canggih dan metodologi yang tepat untuk dapat mengekstraksi informasi yang bermakna dari lautan data yang dihasilkan.

Media sosial seperti Twitter berperan sebagai ruang diskusi publik yang dinamis, tempat beragam pandangan masyarakat muncul, mulai dari kecaman, kekhawatiran, dukungan, hingga analisis dan spekulasi. Percakapan yang terjadi di platform ini meninggalkan jejak digital yang berharga, yang dapat dimanfaatkan untuk memahami bagaimana masyarakat Indonesia merespons dan memproses informasi terkait isu terorisme, serta mengidentifikasi pola interaksi yang mencerminkan opini publik secara kolektif.

Analisis sentimen, sebagai bagian dari *text mining*, berfokus pada pengolahan teks untuk mengidentifikasi emosi, opini, dan persepsi seseorang terhadap suatu topik melalui penerapan teknik pemrosesan bahasa alami. [6]. Di platform media sosial, analisis sentimen dapat memberikan informasi penting dengan cara mengkategorikan komentar pengguna ke dalam kelompok sentimen yang positif, negatif, atau netral dengan efisien dan tepat [7]. Informasi ini tidak hanya mendukung penelitian akademis, tetapi juga memiliki dampak praktis yang besar bagi pengambil kebijakan, lembaga keamanan, dan organisasi kontraterorisme. Dengan memahami respons masyarakat terhadap isu terorisme di media sosial, pihak terkait dapat menyusun strategi komunikasi yang lebih efektif, mengantisipasi potensi radikalisasi, serta merancang program deradikalisasi yang lebih tepat sasaran.

Efektivitas analisis sentimen sangat ditentukan oleh pemilihan algoritma yang tepat sesuai dengan konteks, studi kasus, dan sumber data [8]. Dalam lingkungan data media sosial yang rumit dan tidak teratur, algoritma *machine learning* seperti *Support Vector Machine* (SVM) terbukti memberikan hasil yang lebih baik dibandingkan dengan algoritma lainnya, termasuk *Naïve Bayes*. Penelitian yang dilakukan sebelumnya mengindikasikan bahwa SVM memperoleh tingkat akurasi sebesar 94% ketika dibandingkan dengan *Naïve Bayes* yang memiliki akurasi 91% dalam menilai sentimen tentang topik Ibukota Nusantara [9]. Keunggulan utama dari SVM terletak pada kapasitasnya dalam mengoptimalkan margin pemisahan antar kelas, sehingga meningkatkan akurasi klasifikasi, terutama pada dataset yang bersifat kompleks dan tidak terstruktur.

Penelitian lain yang menggunakan pendekatan *Long Short-Term Memory* (LSTM) pada analisis sentimen terkait isu terorisme di media sosial menunjukkan bahwa akurasi model masih tergolong rendah. Dalam penelitian tersebut, penerapan LSTM dengan jumlah unit tertentu hanya mampu mencapai akurasi tertinggi sebesar 59,34%, meskipun menggunakan *Natural Language Processing* (NLP) untuk pengolahan data [10]. Hal ini mengindikasikan bahwa LSTM mungkin kurang optimal dalam menangani analisis sentimen dengan distribusi data yang tidak merata dan kompleksitas tinggi. Berdasarkan kelemahan tersebut, penelitian ini mencoba menerapkan algoritma SVM sebagai alternatif untuk meningkatkan akurasi klasifikasi sentimen terkait isu terorisme.

Permasalahan yang muncul dalam penelitian ini adalah ketidakseimbangan distribusi data (*imbalanced dataset*) yang umum terjadi dalam analisis sentimen media sosial, khususnya pada topik

sensitif seperti terorisme. Berdasarkan observasi awal terhadap data Twitter terkait terorisme, ditemukan bahwa distribusi sentimen sangat tidak merata, dengan prevalensi sentimen negatif yang secara signifikan mendominasi dibandingkan ekspresi sentimen positif maupun netral. Ketidakseimbangan ini dapat menyebabkan bias dalam model klasifikasi, di mana algoritma cenderung mengklasifikasikan sebagian besar data ke dalam kelas mayoritas, sehingga mengurangi akurasi prediksi untuk kelas minoritas. Sebagai upaya mengatasi permasalahan ketidakseimbangan data, penelitian ini menggunakan teknik *Random Over Sampling* setelah metode tahap pelabelan data. *Random Over Sampling* memilih dan menduplikasi data secara acak dari kelas minoritas pada data training sehingga jumlah data pada kelas tersebut bertambah [11]. Metode ini berperan dalam menyeimbangkan distribusi kelas sehingga dapat mengoptimalkan kinerja model dalam mengklasifikasikan data secara lebih akurat dan konsisten.

Penelitian ini bertujuan untuk mengaplikasikan algoritma *Support Vector Machine* dalam menganalisis sentiment masyarakat di Twitter terkait isu terorisme, dengan mengintegrasikan teknik *Random Over Sampling* guna mengatasi ketidakseimbangan data serta ekstraksi fitur TF-IDF untuk meningkatkan representasi teks dan kualitas data analisis. Model yang dikembangkan diharapkan mampu melakukan klasifikasi opini publik ke dalam kategori positif, negatif, dan netral dengan akurat, sekaligus mengevaluasi efektivitas kombinasi teknik tersebut dalam meningkatkan performa klasifikasi pada dataset yang tidak seimbang dan kompleks.

TINJAUAN PUSTAKA

Penelitian mengenai analisis sentimen pada media sosial telah menjadi salah satu fokus utama dalam bidang pemrosesan teks dan analisis data, khususnya untuk memahami pandangan masyarakat terhadap berbagai isu sosial, politik, dan keamanan, termasuk terorisme. Data yang dihasilkan oleh pengguna media sosial, seperti Twitter, memiliki karakteristik dinamis, masif, dan tidak terstruktur, sehingga memerlukan metode pengolahan yang efektif. Salah satu metode yang paling sering digunakan adalah klasifikasi sentimen dengan memanfaatkan algoritma pembelajaran mesin.

Wulandari et al. [5] membandingkan performa algoritma *Naïve Bayes* dengan metode lainnya dalam analisis sentimen pada Twitter terkait isu ijazah palsu Jokowi. Hasil penelitian mengungkapkan bahwa algoritma SVM mencapai tingkat akurasi yang lebih tinggi, yaitu 69,23%, dibandingkan algoritma *Naïve Bayes* yang hanya mencapai 65%. Oleh karena itu, SVM direkomendasikan sebagai algoritma yang lebih efektif dan reliabel dalam analisis sentimen pada data media sosial, terutama untuk isu-isu yang berpotensi menimbulkan polarisasi opini seperti dugaan ijazah palsu.

Pradana et al. [12] melakukan penelitian yang berfokus pada analisis sentimen terhadap kinerja pemerintahan dengan menggunakan algoritma *Naïve Bayes Classifier* (NBC), *K-Nearest Neighbors* (KNN), dan *Support Vector Machine* (SVM). Data awal yang digunakan berjumlah 20.328 *tweet*. Setelah melalui tahap *preprocessing*, data tersisa sebanyak 15.306 *tweet* yang kemudian diolah dengan pemberian bobot menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Sentimen pada data tersebut diklasifikasikan menjadi positif, negatif, dan netral dengan bantuan VADER. Hasil penelitian mengungkapkan bahwa metode NBC mencapai akurasi sebesar 80,24%, KNN sebesar 75,37%, dan SVM memberikan akurasi tertinggi sebesar 85,47%. Penelitian ini menegaskan pentingnya tahap *preprocessing* dan pembobotan dalam analisis sentimen serta menunjukkan keunggulan metode *SVM* dalam hal akurasi. Azzahra et al. [9] menunjukkan bahwa penggunaan pembobotan TF-IDF yang dikombinasikan dengan algoritma *SVM* secara konsisten menunjukkan performa akurasi analisis sentimen yang lebih unggul, yaitu 94%, dibandingkan dengan *Naïve Bayes* yang mencapai 91%. Hasil ini mengindikasikan bahwa SVM dapat dijadikan sebagai alternatif yang lebih efektif dalam menganalisis sentimen terkait Ibukota Nusantara. Selain itu, penelitian ini juga menekankan pentingnya penggunaan ekstraksi fitur TF-IDF dalam meningkatkan akurasi analisis sentimen.

Penelitian lain yang membandingkan model klasifikasi *Random Forest* dan *Support Vector Machine* dalam analisis sentimen terkait *quick count* pemilu 2024 menggunakan kedua algoritma tersebut untuk mengklasifikasikan data *tweet*. Hasil penilaian menggunakan *confusion matriks* menunjukkan bahwa *Random Forest* memperoleh tingkat akurasi sebesar 78%, sementara SVM memiliki akurasi lebih tinggi yaitu 80%. Dari penelitian ini disimpulkan bahwa SVM lebih efektif dalam mengklasifikasikan *tweet* tentang *quick count* Pemilu 2024 dengan selisih akurasi 2% dibandingkan *Random Forest* [13].

Peneltian yang dilakukan oleh Azzahra et al. [14] terkait analisis sentimen opini masyarakat terhadap vaksinasi *booster COVID-19* mengungkapkan bahwa data Twitter dapat dimanfaatkan secara efektif untuk menggambarkan opini publik. Penelitian ini menemukan bahwa algoritma SVM menunjukkan tingkat akurasi yang paling tinggi dibandingkan dengan algoritma *Naïve Bayes* dan *Decision Tree*, yaitu masing-masing 83,33%, 70,00%, dan 79,00%. Hasil penelitian ini menunjukkan bahwa algoritma SVM memiliki kinerja rata-rata yang lebih baik dibandingkan dengan dua algoritma lainnya dalam analisis sentimen mengenai isu vaksinasi *booster COVID-19*.

Fathoniah dan Rozikin [10] melaksanakan penelitian yang bertujuan untuk mengevaluasi pengaruh penerapan LSTM (*Long Short-Term Memory*) dan NLP (*Natural Language Processing*) dalam analisis sentimen masyarakat terkait isu terorisme di media sosial Twitter. Hasil penelitian menunjukkan bahwa

penggunaan LSTM dengan jumlah unit sebesar 20 mencapai tingkat akurasi tertinggi, yaitu 59,34%, dibandingkan dengan jumlah unit lainnya, seperti 10, 30, 40, 50, 60, 70, 80, 90, dan 100. Namun, tingkat akurasi ini masih tergolong rendah untuk kebutuhan analisis sentimen, terutama pada isu sensitif seperti terorisme. Hal ini menunjukkan bahwa algoritma LSTM belum optimal dalam menangani kompleksitas data sentimen dengan distribusi yang tidak merata, sehingga membutuhkan pendekatan lain untuk meningkatkan akurasi.

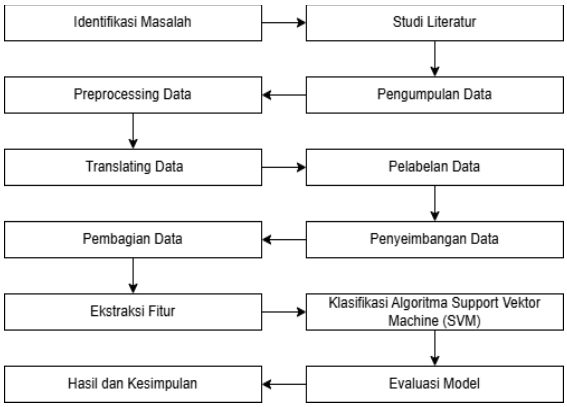
Secara keseluruhan, algoritma SVM sering dianggap lebih sesuai untuk mengolah data teks dari media sosial yang cenderung tidak terstruktur, beragam, dan kaya akan nuansa emosional. Hal ini disebabkan oleh pendekatan margin maksimumnya yang mampu memisahkan kelas sentimen secara lebih efektif. Oleh karena itu, SVM kerap menjadi pilihan utama dalam penelitian yang mengutamakan akurasi klasifikasi, meskipun algoritma ini membutuhkan biaya komputasi yang relatif lebih tinggi.

Meskipun telah banyak studi yang mengulas penerapan algoritma SVM dalam analisis sentimen, masih sangat sedikit penelitian yang secara mendalam membahas isu-isu spesifik seperti terorisme. Terorisme, sebagai topik yang kompleks dan sensitif, membutuhkan pendekatan yang lebih mendalam untuk memahami pola opini publik secara efektif. Selain itu, penggunaan metode seperti *Random Over Sampling* untuk menangani ketidakseimbangan data juga belum banyak diterapkan dalam konteks ini. Padahal, ketidakseimbangan data dapat memengaruhi akurasi model secara signifikan, terutama dalam data dengan distribusi kelas yang tidak merata, seperti sentimen negatif yang cenderung mendominasi diskusi terkait isu terorisme.

Dengan demikian, penelitian ini menawarkan kebaruan (*state of the art*) dengan mengintegrasikan algoritma *Support Vector Machine*, teknik Random Over Sampling, dan ekstraksi fitur TF-IDF dalam analisis sentimen masyarakat Indonesia terkait isu terorisme di *platform* Twitter. Penelitian ini menggunakan data awal yang berbahasa Indonesia, yang kemudian ditranslasi setelah proses *preprocessing* agar dapat digunakan dengan VADER untuk pelabelan sentimen otomatis. Fokus utama penelitian ini adalah pada konteks terorisme dalam media sosial berbahasa Indonesia, yang masih jarang dieksplorasi dalam penelitian sebelumnya. Pendekatan yang diterapkan mencakup penggunaan *Random Over Sampling* untuk mengatasi ketidakseimbangan data serta optimalisasi ekstraksi fitur melalui TF-IDF. Berbeda dengan penelitian terdahulu yang menggunakan LSTM, algoritma SVM dalam penelitian ini diharapkan mampu meningkatkan akurasi klasifikasi sentimen secara signifikan, terutama pada data yang kompleks dan tidak seimbang. Kombinasi metodologi ini diharapkan dapat memberikan kontribusi baru dalam analisis sentimen untuk isu-isu sensitif dan menawarkan wawasan empiris mengenai opini publik Indonesia terhadap ancaman terorisme di media sosial.

METODOLOGI PENELITIAN

Pada Gambar 1 di bawah menampilkan bagan alur yang menggambarkan tahapan penelitian secara sistematis, bertujuan untuk membantu peneliti menjalankan proses penelitian dari awal hingga akhir. Dengan mengikuti bagan tersebut, peneliti dapat memastikan bahwa setiap tahapan penelitian berlangsung secara terintegrasi dan mengikuti skema penelitian yang telah ditetapkan. Hal ini memberikan panduan yang jelas untuk menjaga kelancaran dan keberhasilan penelitian.



Gambar 1. Bagan Alur Penelitian
Sumber : Hasil Olah Data

a. *Identifiksai Masalah*

Fenomena terorisme termasuk dalam kategori isu sensitif yang kerap memicu diskursus publik di berbagai platform media sosial, termasuk Twitter. Ekspresi masyarakat yang beragam dalam bentuk opini, komentar, dan tanggapan dengan perspektif multidimensional menciptakan kompleksitas dalam klasifikasi sentimen publik. Penelitian ini memfokuskan analisis pada polaritas sentimen masyarakat terkait isu terorisme dengan menerapkan algoritma *Support Vector Machine* (SVM), sebuah pendekatan *machine learning* yang telah terbukti efektivitasnya dalam pemrosesan data tekstual. Melalui identifikasi pola dan kecenderungan afektif, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pemahaman komprehensif mengenai dinamika pandangan masyarakat terhadap fenomena terorisme.

b. *Studi Literatur*

Penulis melaksanakan tinjauan literatur dengan mengacu pada berbagai jurnal ilmiah dari penelitian terdahulu yang relevan dengan analisis sentimen. Langkah ini bertujuan untuk memperoleh pemahaman yang komprehensif serta gambaran yang terstruktur terkait penelitian yang akan dilakukan.

c. *Pengumpulan Data*

Proses pengumpulan data adalah salah satu langkah yang sangat penting, yang berpengaruh

besar terhadap mutu dan tingkat ketepatan hasil penelitian. Data yang diperoleh dengan menggunakan metode yang benar akan menjadi dasar yang kuat untuk proses analisis dan pengambilan kesimpulan, sehingga dapat menghasilkan jawaban yang akurat dan relevan terhadap permasalah yang dianalisis [15]. Data dalam penelitian ini adalah data sekunder yang diperoleh dari Kaggle, sebuah platform berbagi dataset yang sering digunakan untuk keperluan analisis data dan pembelajaran mesin. Dataset ini merupakan data mentah yang berisi cuitan di Twitter yang dikumpulkan menggunakan kata kunci "teroris" dalam bahasa Indonesia. Dataset yang digunakan telah melalui proses seleksi untuk memastikan relevansi dan kualitasnya dalam mendukung penelitian ini. Detail lengkap mengenai dataset tersebut dapat diakses melalui pranala berikut: https://www.kaggle.com/datasets/linkgish/indonesian-tweet-about-teroris-terrorism. Dataset ini menyediakan data opini masyarakat yang berkaitan dengan isu terorisme, sehingga memungkinkan analisis sentimen secara komprehensif menggunakan algoritma yang telah dirancang.

d. *Preprocessing Data*

Tahap *preprocessing* merupakan langkah krusial dalam penyiapan data yang berfungsi untuk menyederhanakan proses analisis berikutnya dengan melakukan standarisasi dan pemurnian data awal [16]. Tahapan preprocessing dalam penelitian ini dilakukan melalui beberapa langkah, yaitu *case folding* dan pembersihan teks (*cleaning* text), tokenisasi, normalisasi teks (*normalization* text), penghilangan kata-kata umum yang tidak bermakna (*stopword removal*), serta *stemming* untuk mengubah kata ke bentuk dasarnya.

e. *Translating Data*

Setelah menjalani proses *preprocessing* yang melibatkan beberapa langkah, Data teks yang semula dalam bahasa Indonesia kemudian diubah ke dalam bahasa *Inggris*. Proses penerjemahan ini dilakukan karena alat analisis sentimen yang digunakan, yaitu VADER (*Valence Aware Dictionary and Sentiment Reasoner*), dirancang khusus untuk teks dalam bahasa Inggris. Oleh karena itu, untuk menjamin keakuratan dalam proses pelabelan sentimen, teks yang telah diproses sebelumnya diterjemahkan dengan menggunkan *library* Python GoogleTranslator dari modul deep_translator.

f. *Pelabelan Data*

Setelah data diproses, tahap selanjutnya adalah memberikan label pada data, yaitu mengklasifikasikan setiap data ke dalam kategori sentimen positif, negatif, dan netral. Mengacu pada dataset yang menggunakan bahasa Indonesia, data tersebut perlu diterjemahkan ke dalam bahasa Inggris untuk memungkinkan penggunaan pendekatan berbasis leksikon (*lexicon-based*), yaitu VADER, yang telah terbukti memiliki kinerja lebih baik pada teks dalam bahasa Inggris.

g. *Penyeimbangan Data*

Penyeimbangan data dilakukan dengan menerapkan metode *Random Over Sampling*. Teknik ini bertujuan untuk menambah jumlah data pada kelas minoritas melalui proses duplikasi sampel secara acak. Langkah ini bertujuan untuk menyamakan distribusi data antar kelas, sehingga model klasifikasi dapat mempelajari pola dari masing-masing kelas secara lebih seimbang.

h. *Pembagian Data*

Pada bagian ini, data yang telah dibersihkan dibagi menjadi dua bagian, yaitu data latih dan data uji, dengan pembagian proporsi 80:20. Data pelatihan digunakan untuk mengajarkan model dengan algoritma yang telah ditentukan, agar model dapat memahami pola yang terdapat dalam data. Hasil dari pengolahan data pelatihan ini menjadi landasan untuk melakukan prediksi. Sebaliknya, data uji digunakan untuk mengukur kinerja model serta memverifikasi hasil prediksi yang dihasilkan berdasarkan data pelatihan yang telah diproses sebelumnya.

i. *Ekstraksi Fitur*

Pada tahap ini, karakteristik dari data teks yang telah dibersihkan diambil dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Tahap pembobotan TF-IDF merupakan langkah yang mengonversi data teks menjadi data angka dengan cara menghitung nilai untuk setiap kata atau fitur sesuai dengan frekuensinya [9]. Metode ini dipilih karena dapat memberikan nilai yang mencerminkan pentingnya suatu kata dalam dokumen, sekaligus mengurangi dampak dari kata-kata umum yang sering muncul tetapi tidak sangat berarti, seperti kata penghubung atau kata-kata yang kurang relevan.

j. *Klasifikasi Algoritama Support Vector Machine (SVM)*

Setelah data dibedakan menjadi data pelatihan dan data pengujian, langkah berikutnya adalah melaksanakan klasifikasi dengan memanfaatkan algoritma *Support Vector Machine* (SVM). Algoritma ini dipilih karena memberikan sejumlah kelebihan dalam klasifikasi teks, seperti kemampuannya untuk secara efektif menangani data dengan dimensinya yang tinggi, fleksibilitas dalam memilih fungsi kernel, dan hasil yang akurat pada berbagai penelitian sebelumnya. Di samping itu, Penelitian sebelumnya juga mengindikasikan bahwa SVM seringkali mencapai tingkat akurasi yang lebih tinggi jika dibandingkan dengan metode lainnya, terutama dalam tugas klasifikasi teks.

Dalam studi ini, algoritma SVM diterapkan untuk mengklasifikasikan masalah terorisme ke dalam tiga

kategori sentimen, yaitu positif, negatif, dan netral.Data teks direpresentasikan menggunakan fitur TF-IDF, yang memungkinkan algoritma SVM untuk mengenali pola dalam data secara lebih efektif. Model dilatih menggunakan data pelatihan, sementara evaluasi performa dilakukan pada data pengujian untuk mengukur tingkat akurasi, presisi, dan recall dari prediksi sentimen.

k. *Evaluasi Model*

Confusion matrix digunakan untuk mengevaluasi kinerja model dalam mengklasifikasikan pandangan masyarakat tentang masalah terorisme. Proses evaluasi ini melibatkan berbagai metrik, seperti *accuracy*, *precision*, *recall*, dan *f1-score*, yang memberikan gambaran menyeluruh tentang kualitas prediksi yang dihasilkan oleh model. *Confusion matrix* mencatat jumlah prediksi yang tepat dan tidak tepat untuk setiap kategori sentimen, yaitu positif, negatif, dan netral. Nilai dari indikator-indikator tersebut ditentukan dengan mengacu pada informasi ini.

Selain itu, metode *k-fold cross-validation* digunakan untuk menilai kinerja model secara komprehensif dan *robust*. Dalam metode ini, data dibagi menjadi beberapa lipatan (*fold*), di mana setiap lipatan secara bergantian digunakan sebagai data pelatihan dan data pengujian. Pendekatan ini memastikan bahwa setiap bagian data berperan dalam proses pelatihan dan pengujian model, sehingga kemungkinan terjadinya *overfitting* dapat dikurangi. Selain itu, metode ini memberikan perhitungan yang lebih akurat mengenai kinerja model secara keseluruhan. Hasil evaluasi yang dilakukan melalui *k-fold cross-validation* memberikan kesempatan bagi para peneliti untuk menilai seberapa efektif algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan pandangan masyarakat terhadap isu terorisme.

l. *Hasil dan Kesimpulan*

Tahap terakhir dari penelitian ini menyajikan hasil analisis persepsi publik mengenai isu terorisme menggunakan algoritma *Support Vector Machine* (SVM), yang dinilai melalui metrik akurasi, presisi, recall, dan f1-score dari *confusion matrix*. Penelitian ini juga mengeluarkan kesimpulan mengenai seberapa efektif SVM dalam mengklasifikasikan sentimen dan memberikan saran untuk pengembangan di masa depan. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan teknologi analisis sentimen, khususnya terkait isu-isu sensitif, dengan pendekatan yang menggunakan machine learning yang dapat diandalkan dan valid.

HASIL DAN PEMBAHASAN

a. *Proses Pengumpulan Data*

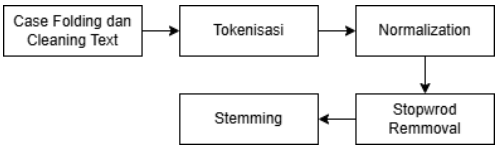
Dataset yang digunakan dalam studi ini diperoleh dari platform Kaggle, berisi kumpulan tweet dari Twitter yang memuat kata kunci "teroris" atau "terorisme" dalam bahasa Indonesia. Dataset ini relevan dengan isu yang menjadi fokus penelitian. Secara keseluruhan, dataset mencakup 210.251 entri, namun peneliti menggunakan sampel acak sebanyak 5.000 data untuk analisis. Pendekatan ini sejalan dengan penelitian Alvian et al. [17], yang menggunakan 5.000 data dalam analisis mereka dan mencapai tingkat akurasi yang cukup tinggi, yaitu sebesar 89%. Sebelum dilakukan analisis sentimen, data tersebut melalui proses *preprocessing* untuk memastikan kualitas dan konsistensinya. Tabel 1 menyajikan contoh tweet yang diambil dari dataset untuk memberikan gambaran tentang data yang digunakan.

Tabel 1. Sampel Tweet Text

Tweet Text
@henrysubiakto Kebenaran itu apakah saat tau ada Ulama teroris kita malah diam dan membiarkannya.
@RANGERmounts Bukan oposisi lgi,mungkin pendukung teroris.
@Pai__C1 Yang membela teroris pokoknya harus diwaspada i

b. *Preprocessing Data*

Preprocessing adalah tahap yang krusial dalam analisis data teks, yang bertujuan untuk membersihkan dan menyiapkan data mentah sehingga siap digunakan oleh algoritma klasifikasi. Pada data media sosial seperti Twitter, proses ini menjadi sangat krusial karena data seringkali mengandung elemen-elemen tidak relevan, seperti simbol, tagar, tautan, *mention*, serta kata-kata slang atau tidak baku. Langkah *preprocessing* tidak hanya mengurangi kompleksitas data, tetapi juga berkontribusi pada peningkatan akurasi model klasifikasi, karena algoritma dapat lebih mudah mengidentifikasi pola dan makna dalam teks yang telah dibersihkan [5]. Dalam penelitian ini, preprocessing diterapkan pada 5.000 data yang diambil sebagai sampel, dengan hanya menggunakan fitur atau kolom yang memuat teks tweet. Namun, setelah melalui proses *preprocessing*, beberapa data ditemukan tidak memiliki isi teks (kosong). Akibatnya, jumlah data yang tersisa setelah preprocessing adalah 4.947, karena data kosong tersebut dihapus untuk menjaga kualitas analisis. Langkah-langkah dalam proses *preprocessing* data dapat dilihat pada ilustrasi berikut.



Gambar 2. Tahapan Preprocessing
Sumber : Hasil Olah Data

- *Case Folding dan Cleaning Text*
  Case folding adalah suatu proses yang bertujuan untuk menyamakan format huruf dengan cara mengubah semua huruf kapital menjadi huruf

kecil. Sementara itu, *cleaning* teks adalah tahap untuk menghilangkan kata-kata atau unsur yang tidak berkaitan dari dokumen, sehingga dapat mengurangi gangguan (*noise*). Elemen yang dihilangkan meliputi karakter *HTML*, kata-kata yang tidak diperlukan, emotikon, tagar (*hashtag*) yang diawali dengan tanda #, *mention username* yang diawali dengan tanda @, angka, serta *URL* (*Uniform Resource Locator*) [18]. Hasil data setelah melalui proses *case folding* dan *cleaning* text disajikan pada Tabel 2.

Tabel 2. Hasil Case Folding dan Cleaning Text

Tweet Text	Hasil Case Folding dan Cleaning Text
TERORIS GAK SIH? oh atau takut menunjukan behel mengerikan itu? https://t.co/0iEHAH2HzJ	teroris gak sih oh atau takut menunjukan behel mengerikan itu
@fotodakwah Buku barang bukti penangkapan teroris?	buku barang bukti penangkapan teroris
@OposisiCerdas Takut lah lawan teroris beneran mah..	takut lah lawan teroris beneran mah

- *Tokenisasi*
  Tokenisasi merupakan proses yang bertujuan untuk memecah sekumpulan kata dalam sebuah kalimat, paragraf, atau dokumen menjadi unit-unit kecil yang disebut token atau istilah kata, yang dapat berdiri sendiri. [19]. Tabel 3 menyajikan hasil dari proses *tokenisasi*.

Tabel 3. Hasil Tokenisasi

Hasil Case Folding dan Cleaning Text	Hasil Tokenisasi
iya tolong jangan jadi teroris ya	['iya', 'tolong', 'jangan', 'jadi', 'teroris', 'ya']
terduga teroris di batam masih diperiksa intensif	['terduga', 'teroris', 'di', 'batam', 'masih', 'diperiksa', 'intensif']
banyak teroris yg masuk pada organisasi masa dan lsm untuk bersembunyi dan menyerang pemerintah yg sah	['banyak', 'teroris', 'yg', 'masuk', 'pada', 'organisasi', 'masa', 'dan', 'lsm', 'untuk', 'bersembunyi', 'dan', 'menyerang', 'pemerintah', 'yg', 'sah']

- *Normalization*
  Pada fase ini, dilakukan langkah untuk menyesuaikan setiap kata yang mengandung elemen tidak standar atau gangguan menjadi kata standar yang siap untuk diproses. Kata-kata yang tidak formal tersebut mencakup istilah-istilah dari bahasa daerah yang tidak sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) serta akronim-akronim yang sering dipakai di media sosial. Sebagai contoh, kata singkatan seperti "tdk" akan dinormalisasi menjadi "tidak" melalui proses ini [20]. Hasil dari proses ini ditampilkan pada Tabel 4 di bawah ini :

Tabel 4. Hasil Normalization

Hasil Tokenisasi	Hasil Normalization
['teroris', 'kok', 'ulama', 'orang', 'bermasalah', 'kok', 'ulama', 'kalau', 'ulama', 'itu', 'panutan', 'ahlak', 'dn', 'perilaku', 'seperti', 'nabi', 'bukan', 'seperti', 'khowarij']	['teroris', 'mengapa', 'ulama', 'orang', 'bermasalah', 'mengapa', 'ulama', 'kalau', 'ulama', 'itu', 'panutan', 'akhlak', 'dan', 'perilaku', 'seperti', 'nabi', 'bukan', 'seperti', 'khowarij']
['satu', 'demi', 'satu', 'gugur', 'oleh', 'para', 'teroris', 'yg', 'jelas', 'mengancam', 'kedaulatan', 'negara']	['satu', 'demi', 'satu', 'gugur', 'oleh', 'para', 'teroris', 'yang', 'jelas', 'mengancam', 'kedaulatan', 'negara']
['kelompok', 'tersebut', 'teroris', 'yg', 'disini', 'bukan', 'musuh']	['kelompok', 'tersebut', 'teroris', 'yang', 'disini', 'bukan', 'musuh']

- *Stopword Removal*
  Stopword removal merupakan langkah penyaringan atau pemilihan kata dalam teks untuk menyingkirkan kata-kata yang tidak memiliki arti penting, seperti kata penghubung, kata keterangan, dan elemen sejenis lainnya, yang tidak diperlukan dalam pemodelan data [21]. Hasil dari proses *stopword removal* dapat dilihat pada Tabel 5.

Tabel 5. Hasil Stopword Removal

Hasil Normalization	Hasil Stopword Removal
['yang', 'teroris', 'saat', 'ini', 'rezim', 'ini']	['teroris', 'rezim']
['mulai', 'menggiring', 'opini', 'padahal', 'pantas', 'dihukum', 'lebih', 'berat', 'dari', 'hukuman', 'mati', 'bagi', 'aktor', 'intelektual', 'teroris']	['menggiring', 'opini', 'dihukum', 'berat', 'hukuman', 'mati', 'aktor', 'intelektual', 'teroris']
['hanya', 'kaum', 'dungu', 'yang', 'percaya', 'ada', 'teroris']	['kaum', 'dungu', 'percaya', 'teroris']

- *Stemming*
  Proses *stemming* adalah langkah terakhir dalam preprocessing pada penelitian ini. Tahap ini

bertujuan untuk mengubah kata-kata dalam dokumen menjadi bentuk dasar atau akar kata. Tujuan utama dari proses *stemming* adalah untuk menyederhanakan berbagai bentuk kata yang muncul akibat penggunaan afiks, sehingga kata-kata dengan makna dasar yang serupa dapat dikelompokkan bersama [22]. Proses stemming kata dalam dokumen yang menggunakan bahasa Indonesia cukup rumit, karena mencakup penghilangan semua afiks yang terdapat pada kata-kata dalam tweet. Pada penelitian ini, proses stemming dilakukan menggunakan *library Sastrawi* yang dirancang khusus untuk Bahasa Indonesia [23]. Tabel 6 menampilkan hasil dari proses ini.

Tabel 6. Hasil Stemming

Hasil Stopword Removal	Hasil Stemming
['teroris', 'dipelihara', 'menggelontorkan', 'anggaran', 'semoga', 'opini', 'salah']	['teroris', 'pelihara', 'gelontor', 'anggar', 'moga', 'opini', 'salah']
['orang', 'berbahaya', 'kerukunan', 'umat', 'beragama', 'terlibat', 'jaringan', 'teroris']	['orang', 'bahaya', 'rukun', 'umat', 'agama', 'libat', 'jaring', 'teroris']
['aparat', 'keamanan', 'aman', 'masyarakat', 'ancaman', 'kebiadaban', 'teroris', 'kkb', 'menyengsarakan', 'masyarakat', 'papua']	['aparat', 'aman', 'aman', 'masyarakat', 'ancam', 'biadab', 'teroris', 'kkb', 'sengsara', 'masyarakat', 'papua']

c. *Translating Data*

Setelah data teks dalam bahasa Indonesia menjalani proses *preprocessing*, langkah berikutnya adalah menerjemahkan teks tersebut ke dalam bahasa Inggris menggunakan program *library Python GoogleTranslator* dari modul *deep_translator*. Proses translasi ini diperlukan karena alat analisis sentimen yang digunakan, yaitu *VADER*, hanya dirancang untuk memproses teks berbahasa Inggris sehingga memerlukan data dalam bahasa Inggris agar hasilnya akurat.

Dengan penerjemahan data ke dalam bahasa Inggris, proses analisis sentimen menggunakan VADER menjadi lebih efisien, karena kamus dan algoritma VADER lebih komprehensif dan sah untuk teks dalam bahasa Inggris. Hal ini memungkinkan model untuk mengenali perbedaan dalam sentimen dengan lebih akurat, sehingga meningkatkan ketepatan klasifikasi sentimen dalam penelitian ini. Hasil dari proses ini dapat dilihat pada Tabel 7.

Tabel 7. Hasil Translating Data

Hasil Translating Data
Papuan people hate KKB terrorists to spread the provocation of killing the Indonesian National Police Protection of the Papuan community
How do Papuan terrorists feel sorry for the soldiers only sacrifice
keep the community of the boundary fear of attack terrorists

d. *Pelabelan Data*

Setelah proses penerjemahan data selesai, langkah selanjutnya adalah pelabelan data, yang berarti mengelompokkan sentimen dari setiap komentar pengguna Twitter. Proses pelabelan ini menggunakan pendekatan berbasis leksikon (*lexicon-based*) dengan bantuan alat VADER (*Valence Aware Dictionary and Sentiment Reasoner*), yang dirancang untuk teks berbahasa Inggris. Data yang telah diterjemahkan sebelumnya memastikan akurasi dan konsistensi dalam pengklasifikasian sentimen. mengindikasikan bahwa 4.398 data termasuk dalam kategori negatif, 368 data dalam kategori positif, dan 181 data berada dalam kategori netral. Contoh hasil pelabelan data untuk sentimen positif, negatif, dan netral dapat dilihat pada Tabel 8.

Tabel 8. Hasil Pelabelan Data

Text	Score	Label
I hope the KKB peace is tired quickly, sorry	0.4939	Positif
Build a special prison for a safe terrorist to the maximum center of deradicalization	-0.5267	Negatif
quotes teroris	0	Netral

e. *Penyeimbangan Data*

Setelah proses pelabelan data dilakukan, analisis distribusi data menunjukkan terdapatnya ketidakseimbangan jumlah antara kelas-kelas sentimen. Kelas negatif memiliki jumlah data yang secara signifikan lebih banyak dibandingkan dengan kelas positif dan kelas netral, dengan rincian 4.398 data untuk sentimen negatif, 368 data untuk sentimen positif, dan 181 data untuk sentimen netral. Ketidakseimbangan ini dapat menyebabkan bias dalam pelatihan model klasifikasi, di mana algoritma cenderung memprioritaskan kelas mayoritas, sehingga mengurangi akurasi pada kelas minoritas.

Untuk mengatasi masalah ini, metode *Random Over Sampling* diterapkan guna menyeimbangkan distribusi data antar kelas. *Random Over Sampling* adalah teknik untuk menyeimbangkan data dengan menduplikasi sampel pada kelas minoritas secara acak [24]. Pemilihan metode ini didukung oleh hasil penelitian Kurniadi et al. [25], yang menunjukkan bahwa *resampling* menggunakan *Random Over Sampling* menghasilkan akurasi yang lebih tinggi, yaitu sebesar 99,8%, dibandingkan dengan metode lainnya seperti *Random Under Sampling* (97,2%), SMOTE (97,4%), dan Tomek Links (98,7%). Setelah menggunakan metode ini, jumlah data di setiap

kategori menjadi seimbang, di mana setiap kategori memiliki 4.398 data. Proses ini memastikan model dapat belajar secara adil tanpa bias terhadap kelas mayoritas.

f. Pembagian Data

Hasil pembagian data menunjukkan bahwa jumlah data latih sebanyak 10.555 sampel, sementara jumlah data uji sebanyak 2.639 sampel, dengan proporsi pembagian 80:20. Pembagian ini bertujuan untuk menjamin bahwa model menerima jumlah data yang cukup untuk proses pembelajaran serta data yang memadai untuk keperluan pengujian, sehingga evaluasi performa model dapat dilakukan secara menyeluruh.

g. Ekstraksi Fitur

Setelah proses pembagian data selesai, langkah berikutnya adalah ekstraksi fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Proses ini bertujuan untuk mengonversi data teks menjadi angka dengan menghitung nilai setiap kata berdasarkan frekuensi kemunculan kata tersebut dalam dokumen tertentu serta seberapa jarang kata itu muncul di seluruh dokumen. Baris kode yang menunjukkan penerapan TF-IDF dapat dilihat pada Gambar 3 di bawah ini.

```
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf_vectorizer = TfidfVectorizer(max_features=5000, ngram_range=(1, 2))
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)
```

Gambar 3. Ekstraksi Fitur
Sumber : Hasil Olah Data

Kode implementasi TF-IDF yang ditampilkan pada Gambar 5 menunjukkan langkah-langkah untuk mengubah data teks menjadi representasi numerik. Pustaka TfidfVectorizer dari scikit-learn digunakan dengan parameter max_features=5000 untuk membatasi jumlah fitur hingga 5.000 kata, dan ngram_range=(1, 2) untuk mempertimbangkan unigram dan bigram. Data latih (X_train) diubah menjadi bentuk matriks TF-IDF dengan menggunakan metode fit_transform, sementara data uji (X_test) diubah dengan penerapan metode transform. Pendekatan ini menjamin bahwa fitur-fitur yang penting dapat dikenali dengan lebih efisien.

h. Implementasi Algoritma Support Vector Machine (SVM)

Implementasi algoritma *Support Vector Machine* (SVM) dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python. Python menyediakan pustaka scikit-learn yang mendukung implementasi SVM dengan berbagai metode dan algoritma untuk pengembangan *machine learning*. Scikit-learn merupakan pustaka yang komprehensif dan sering digunakan untuk tugas-tugas klasifikasi, termasuk SVM. Baris kode untuk penerapan SVM dapat dilihat pada Gambar 4 di bawah ini.

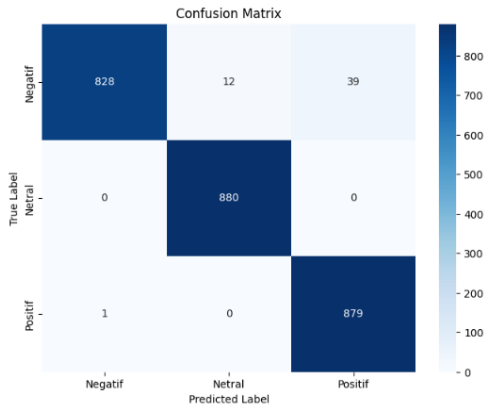
```
from sklearn.svm import LinearSVC
svm_clf = LinearSVC(random_state=42)
svm_clf.fit(X_train_tfidf, y_train)
```

Gambar 4. Kode Implemenatsi Algoritma SVM
Sumber : Hasil Olah Data

Kode pada Gambar 6 mengimplementasikan algoritma *Support Vector Machine* (SVM) dengan memanfaatkan kelas LinearSVC dari pustaka scikit-learn. *Linear Support Vector Classification* (LinearSVC) menggunakan metode *Support Vector Machine* untuk memprediksi kelas dari titik data input melalui identifikasi *hyperplane* optimal yang memisahkan dua kelas atau lebih [26]. Algoritma ini sangat akurat dan telah terbukti efektif untuk tugas klasifikasi biner (dua kelas) maupun *multiclass* (lebih dari dua kelas). LinearSVC memanfaatkan kernel linear yang memproyeksikan data ke ruang berdimensi tinggi tanpa memerlukan transformasi non-linear, sehingga sangat sesuai untuk menangani data yang dapat dipisahkan secara linear atau mendekati linear [26]. Model ini dibuat dengan pengaturan random_state untuk menjaga konsistensi hasil dan dilatih menggunakan data fitur TF-IDF (X_train_tfidf) beserta labelnya (y_train) melalui metode fit(). Proses pelatihan ini memungkinkan model mempelajari pola dari data latih guna melakukan klasifikasi sentimen secara efektif. LinearSVC sangat efisien dalam menangani dataset besar dan sering digunakan dalam tugas-tugas berbasis teks seperti penelitian ini.

i. Evaluasi Model

Evaluasi kinerja model *Support Vector Machine* (SVM) dilaksanakan menggunakan *confusion matrix* yang merepresentasikan jumlah prediksi yang tepat dan tidak tepat untuk setiap kategori sentimen: negatif, netral, dan positif. Berdasarkan *confusion matrix*, untuk sentimen negatif, dari 879 data aktual, model berhasil memprediksi 828 data dengan akurat, sedangkan 12 data keliru diprediksi sebagai netral, dan 39 lainnya salah diprediksi sebagai positif. Pada kategori netral, semua dari 880 data aktual diprediksi dengan benar tanpa kesalahan. Sedangkan pada kategori positif, dari 880 data aktual, sebanyak 879 prediksi benar, dan hanya 1 data salah diprediksi sebagai negatif. Hasil dari *confusion matrix* dapat diamati pada Gambar 5 di bawah ini.



Gambar 5. Visualisasi Confusion Matrix
Sumber : Hasil Olah Data

Model ini menunjukkan kinerja yang sangat baik dengan tingkat *accuracy* mencapai 98,02%. Selain itu, angka *precision* mencapai 98,09%, angka *recall* sebesar 98,02%, dan nilai *f1-score* sebesar 98,01% memperkuat argumen bahwa model ini dapat memberikan prediksi yang tidak hanya tepat, tetapi juga konsisten dalam mengenali setiap kategori sentimen. Hasil dari uji coba dapat diperhatikan pada gambar 9.

Tabel 9. Evaluasi Model

Acucuracy	Precision	Recall	F1-Score
98.02%	98.09%	98.02%	98.01%

Untuk mengevaluasi performa model secara lebih menyeluruh, digunakan metode *k-fold cross-validation* dengan nilai k sebesar 5. *K-fold cross-validation* merupakan teknik evaluasi model yang bertujuan untuk mengukur performa secara lebih robust dan menyeluruh. Dalam metode ini, dataset dibagi menjadi k lipatan (*fold*), di mana setiap lipatan secara bergantian digunakan sebagai data pengujian sementara lipatan lainnya digunakan sebagai data pelatihan [27]. Proses ini memastikan bahwa setiap informasi dalam kumpulan data digunakan dengan baik sebagai data pelatihan maupun data pengujian, sehingga mengurangi kemungkinan terjadinya bias akibat pembagian data yang tidak seimbang. Hasil evaluasi *k-fold cross-validation* ditampilkan pada Tabel 10, yang mencakup metrik *accuracy*, *precision*, *recall*, dan *f1-score* untuk setiap lipatan, serta nilai rata-rata dan standar deviasi untuk memberikan gambaran performa model secara keseluruhan.

Tabel 10. Hasil *K-Fold Cross-Validation*

Fold	Accuracy	Precision	Recall	F1-Score
1	97.73%	97.82 %	97.73%	97.71%
2	97.35%	97.47%	97.35%	97.33%
3	97.99%	98.07%	97.99%	97.98%
4	97.95%	98.02%	97.95%	97.95%
5	97.84%	97.92%	97.84%	97.83%
Rata-Rata	97.77%	97.86%	97.77%	97.76%

Hasil ini mengindikasikan bahwa model menunjukkan kinerja yang stabil dan konsisten di setiap lipatan data. Dengan rata-rata *accuracy* sebesar 97,77%, *precision* sebesar 97,86%, *recall* sebesar 97,77%, dan *f1-score* sebesar 97,76%, model ini dapat dipercaya dalam mengidentifikasi setiap kategori sentimen dengan tingkat kesalahan yang minimal. Nilai-nilai ini menunjukkan kemampuan model dalam memberikan prediksi yang tepat dan dapat dipercaya dalam berbagai situasi pengujian.

KESIMPULAN DAN SARAN

Penelitian ini berhasil menunjukkan bahwa algoritma *Support Vector Machine* (SVM) merupakan metode yang efektif dalam mengelompokkan sentimen masyarakat terkait isu terorisme. Proses penelitian dimulai dengan *preprocessing* teks dalam bahasa Indonesia, yang kemudian diterjemahkan ke dalam bahasa Inggris untuk diberikan label menggunakan VADER. Untuk mengatasi ketidakseimbangan data, digunakan metode *Random Over Sampling*, sedangkan ekstraksi fitur dilakukan menggunakan TF-IDF untuk menghasilkan representasi teks yang sesuai bagi model. Evaluasi dengan menggunakan *confusion matrix* menunjukkan bahwa model memperoleh *accuracy* sebesar 98,02%, *precision* sebesar 98,09%, *recall* sebesar 98,02%, dan *f1-score* sebesar 98,01%. Sementara hasil evaluasi *k-fold cross-validation* dengan nilai k sebesar 5, menunjukkan bahwa model memiliki performa yang stabil dan konsisten pada seluruh lipatan data. Dengan rata-rata *accuracy* sebesar 97,77%, *precision* sebesar 97,86%, *recall* sebesar 97,77%, dan *f1-score* sebesar 97,76%, model dapat diandalkan untuk mengenali setiap kategori sentimen dengan tingkat kesalahan yang rendah. Hasil ini menunjukkan bahwa proses *preprocessing*, penyeimbangan data, dan ekstraksi fitur yang dilaksanakan sangat berkontribusi terhadap kinerja tinggi model. Namun, kelemahan kecil ditemukan dalam klasifikasi data negatif dan positif, dengan beberapa kesalahan prediksi yang masih terjadi.

Keterbatasan utama dalam penelitian ini terletak pada penggunaan VADER sebagai alat pelabelan sentimen, yang dirancang untuk Bahasa *Inggris*. Oleh karena itu, menerjemahkan data dari Bahasa Indonesia dapat berisiko menyebabkan distorsi makna dan bias dalam anotasi. Agar akurasi pelabelan dapat ditingkatkan, disarankan untuk menggunakan model klasifikasi yang telah dilatih pada kumpulan data Bahasa Indonesia atau melakukan pelabelan secara manual dengan validasi dari ahli. Selain itu, perlu dilakukan pengujian terhadap teknik penyeimbangan data seperti SMOTE atau ADASYN sebagai perbandingan dengan *Random Over Sampling*. Penggunaan model yang berbasis Transformer seperti IndoBERT juga dianjurkan karena kemampuan model ini dalam memahami konteks semantik dalam Bahasa Indonesia dengan lebih

mendalam dibandingkan dengan model-model tradisional. Mengingat bahwa data yang digunakan hanya berasal dari Twitter dan telah melalui proses penerjemahan, temuan dari penelitian ini belum dapat diterapkan secara luas. Oleh karena itu, sangat disarankan untuk mengumpulkan data dari berbagai platform serta melakukan analisis dalam bahasa aslinya untuk pengembangan lebih lanjut.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada empat rekan penulis yang telah memberikan kontribusi signifikan dalam pelaksanaan dan penyelesaian penelitian ini. Kerjasama, komitmen, dan pemeriksaan yang teliti dari setiap penulis sangat berperan dalam mencapai tujuan penelitian.

DAFTAR PUSTAKA

[1] Y. Fitriani, "PEMANFAATAN MEDIA SOSIAL SEBAGAI MEDIA PENYAJIAN KONTEN EDUKASI ATAU PEMBELAJARAN DIGITAL," *Journal of Information System, Applied, Management, Accounting and Research*, vol. 5, no. 4, pp. 1006–1013, 2021, doi: 10.52362/jisamar.v5i4.609.

[2] R. D. Pebrianti, "ANALISIS SENTIMEN MASYARAKAT PLATFORM X TERHADAP KORUPSI PT. PERTAMINA (PERSERO) MENGGUNAKAN METODE SVM," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6399.

[3] M. Zulfikar and Aminah, "PERAN BADAN NASIONAL PENANGGULANGAN TERORISME DALAM PEMBERANTASAN TERORISME DI INDONESIA," *Jurnal Pembangunan Hukum Indonesia*, vol. 2, no. 1, pp. 129–144, Jan. 2020, doi: 10.14710/jphi.v2i1.129-144.

[4] R. Pradana and J. Setiyono, "Peran Pendidikan Pancasila Terhadap Pencegahan Penyebaran Terorisme Di Kalangan Pelajar," *Jurnal Pembangunan Hukum Indonesia*, vol. 3, no. 2, pp. 136–154, May 2021, doi: 10.14710/jphi.v3i2.136-154.

[5] O. M. Wulandari, I. Maulana, F. Syamsudin, and R. Waluyo, "PERBANDINGAN ALGORITMA NAIVE BAYES DAN SVM DALAM ANALISIS SENTIMEN TWITTER TERHADAP ISU IJAZAH JOKOWI PALSU," *Jurnal Manajemen Informatika, Sistem Informasi dan Teknologi Komputer (JUMISTIK)*, vol. 4, no. 1, pp. 392–400, 2025, doi: 10.70247/jumistik.v4i1.145.

[6] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, pp. 823–848, Jan. 2023, doi: 10.37905/jjeee.v5i1.15352.

[7] P. A. Permatasari, Linawati, and L. Jasa, "Survei Tentang Analisis Sentimen Pada Media Sosial," *Majalah Ilmiah Teknologi Elektro*, vol. 20, no. 2, p. 177, Dec. 2021, doi: 10.24843/mite.2021.v20i02.p01.

[8] Y. A. Singgalen, "Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Sistematic Literature Review," *Journal of Information Systems and Informatics*, vol. 3, no. 2, pp. 278–302, Jun. 2021, doi: 10.33557/journalisi.v3i2.125.

[9] A. Supian, B. Tri Revaldo, N. Marhadi, L. Efrizoni, and Rahmaddeni, "PERBANDINGAN KINERJA NAÏVE BAYES DAN SVM PADA ANALISIS SENTIMEN TWITTER IBUKOTA NUSANTARA," *JURNAL ILMIAH INFORMATIKA*, vol. 12, no. 01, pp. 15–21, Mar. 2024, doi: 10.33884/jif.v12i01.8721.

[10] S. Fathoniah and C. Rozikin, "Analisis Sentimen Masyarakat terhadap Teroris dalam Media Sosial Twitter menggunakan NLP," *Jurnal Ilmiah Wahana Pendidikan*, vol. 2022, no. 13, pp. 412–419, Aug. 2022, doi: 10.5281/zenodo.6962682.

[11] A. R. D. Saputra and B. Prasetiyo, "Sentiment Analysis on Twitter Social Media Regarding Covid-19 Vaccination with Naive Bayes Classifier (NBC) and Bidirectional Encoder Representations from Transformers (BERT)," *Recursive Journal of Informatics*, vol. 2, no. 2, 2024, doi: 10.15294/rji.v8i1.25356.

[12] H. Y. Pradana, I. Slamet, and E. Zukhronah, "ANALISIS SENTIMEN KINERJA PEMERINTAHAN MENGGUNAKAN ALGORITMA NBC, KNN, DAN SVM," *Prosiding Simposium Nasional Multidisiplin (SinaMu)*, vol. 4, p. 114, Feb. 2023, doi: 10.31000/sinamu.v4i1.7869.

[13] I. Septiana and D. Alita, "Perbandingan Random Forest dan SVM dalam Analisis Sentimen Quick Count Pemilu 2024," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 9, no. 3, pp. 224–233, Dec. 2024, doi: 10.30591/jpit.v9i3.6640.

[14] R. T. Aldisa and P. Maulana, "Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, pp. 106–109, Jun. 2022, doi: 10.47065/bits.v4i1.1581.

[15] S. Romdona, S. S. Junista, and A. Gunawan, "TEKNIK PENGUMPULAN DATA: OBSERVASI, WAWANCARA DAN KUESIONER," *JISOSEPOL: Jurnal Ilmu Sosial Ekonomi dan Politik*, vol. 3, no. 1, pp. 39–47, Jan. 2025, doi: 10.61787/taceee75.

[16] F. M. Sarimole and Kudrat, "Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 783–790, 2024, doi: 10.55338/saintek.v5i1.2702.

[17] V. Alviani, S. Alam, and I. Kurniawan, "ANALISIS SENTIMEN REVIEW APLIKASI WETV PADA PLATFORM TWITTER MENGGUNAKAN SUPPORT

VECTOR MACHINE," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 143–149, Aug. 2023, doi: 10.55123/storage.v2i3.2351.

[18] 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, Nov. 2020, doi: 10.21107/edutic.v7i1.8779.

[19] D. Darwis, N. Siskawati, and Z. Abidin, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL," *Jurnal Tekno Kompak*, vol. 15, no. 1, p. 131, Feb. 2021, doi: 10.33365/jtk.v15i1.744.

[20] J. A. Septian, T. M. Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *Journal of Intelligent System and Computation*, vol. 1, no. 1, pp. 43–49, Aug. 2019, doi: 10.52985/insyst.v1i1.36.

[21] D. D. Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.

[22] A. Pratama, J. Nababan, and N. S. Salsabila, "Algoritma Smith-Waterman Untuk Mengidentifikasi Kemiripan Judul Proyek Mahasiswa," *JIKTEKS : Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 03, no. 01, pp. 22–34, 2024, doi: 10.70404/jikteks.v3i01.125.

[23] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," 2019. [Online]. Available: http://j-ptiik.ub.ac.id

[24] S. Diantika, "PENERAPAN TEKNIK RANDOM OVERSAMPLING UNTUK MENGATASI IMBALANCE CLASS DALAM KLASIFIKASI WEBSITE PHISHING MENGGUNAKAN ALGORITMA LIGHTGBM," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 19–25, Jan. 2023, doi: 10.36040/jati.v7i1.6006.

[25] D. Kurniadi, F. Nuraeni, N. Faturrohman, and A. Mulyani, "Klasifikasi Perputaran Karyawan Perusahaan Menggunakan Algoritma Random Forest dan Random Over-sampling," *Edu Komputika Journal*, vol. 10, no. 2, pp. 93–103, Jul. 2024, doi: 10.15294/edukomputika.v10i2.73782.

[26] T. Hariguna and A. Ruangkanjanases, "Adaptive sentiment analysis using multioutput classification: a performance comparison," *PeerJ Comput Sci*, vol. 9, p. e1378, May 2023, doi: 10.7717/peerj-cs.1378.

[27] P. Arsi and R. Waluyo, "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 8, no. 1, pp. 147–156, Feb. 2021, doi: 10.25126/jtiik.202183944.