JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.219-226

PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE UNTUK MENDETEKSI UJARAN KEBENCIAN DALAM MEDIA SOSIAL TWITTER

Karsten Jonatthan Shallom*[1)], Hendry[2)]
1. Satya Wacana Christian University, Indonesia
2. Satya Wacana Christian University, Indonesia

Article Info
Kata Kunci: Klasifikasi Teks; Media Sosial; Natural Language Processing, Support Vector Machine, Ujaran Kebencian.

Keywords: Hate Speech; Machine Learning; Natural Language Processing; Social Media; Support Vector Machine; Text Classification.

Article history:
Received 27 February 2025
Revised 8 April 2025
Accepted 15 May 2025
Available online 1 March 2026

DOI :
https://doi.org/10.29100/jipi.v11i1.7553

* Corresponding author.
Corresponding Author
E-mail address:
iamkarsten15@gmail.com

ABSTRAK
Penelitian ini mengeksplorasi penerapan algoritma Support Vector Machine untuk mendeteksi ujaran kebencian di platform media sosial Twitter, khususnya dalam konteks bahasa Indonesia. Dengan lebih dari 330 juta pengguna, Twitter menjadi sarana yang rentan terhadap penyebaran ujaran kebencian yang dapat menimbulkan dampak negatif. Tujuan utama dari penelitian ini adalah mengembangkan sistem otomatis yang mampu mengidentifikasi ujaran kebencian secara efektif. Dataset yang digunakan terdiri dari 1564 tweet berbahasa Indonesia yang diambil dari isu politik pada tahun 2021. Proses analisis meliputi langkah-langkah seperti tokenisasi, stemming, dan penandaan kelas kata, diikuti dengan klasifikasi menggunakan SVM. Hasil penelitian menunjukkan bahwa 92.8% dari tweet yang dianalisis termasuk dalam kategori "no hate speech," sementara 7.2% teridentifikasi sebagai "hate speech." Model SVM menunjukkan performa yang sangat baik dengan akurasi mencapai 97.19%, recall 97.19%, presisi 97.28%, dan F1 Score 96.82%, tanpa adanya False Negatives. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam menciptakan lingkungan online yang lebih aman dan positif, serta meningkatkan pemahaman tentang karakteristik bahasa Indonesia dalam konteks deteksi ujaran kebencian.

ABSTRACT
This study explores the application of the Support Vector Machine (SVM) algorithm for detecting hate speech on the social media platform Twitter, specifically in the context of the Indonesian language. With over 330 million users, Twitter is a platform susceptible to the spread of hate speech, which can have negative consequences. The primary objective of this research is to develop an automated system capable of effectively identifying hate speech. The dataset used consists of 1,564 Indonesian tweets collected during 2021 related to political issues. The analysis process includes steps such as tokenization, stemming, and part-of-speech tagging, followed by classification using SVM. The results indicate that 92.8% of the analyzed tweets fall into the "no hate speech" category, while 7.2% are identified as "hate speech." The SVM model demonstrates excellent performance with an accuracy of 97.19%, recall of 97.19%, precision of 97.28%, and an F1 Score of 96.82%, with no False Negatives. This research is expected to significantly contribute to creating a safer and more positive online environment and enhance the understanding of the characteristics of the Indonesian language in the context of hate speech detection.

I. PENDAHULUAN

Era digital telah membawa perubahan besar dalam cara manusia berkomunikasi dan berinteraksi. Media sosial, khususnya Twitter (sekarang dikenal sebagai X), telah menjadi platform yang sangat berpengaruh dalam pertukaran informasi dan pendapat. Dengan lebih dari 330 juta pengguna aktif di seluruh dunia, Twitter menjadi salah satu media sosial terpopuler, termasuk di Indonesia, di mana pengguna Twitter mencapai lebih dari 14 juta orang. Namun, seiring dengan meningkatnya popularitas dan aksesibilitas Twitter, platform ini juga menjadi sarana penyebaran ujaran kebencian.

Namun, seiring dengan meningkatnya popularitas dan aksesibilitas Twitter, media sosial ini juga menjadi wadah bagi penyebaran ujaran kebencian. Ujaran kebencian dapat diartikan sebagai bentuk ekspresi yang mendorong, mempromosikan, atau membenarkan kebencian terhadap kelompok tertentu berdasarkan ras, agama, etnis, gender, atau faktor lainnya. Di Indonesia, perbedaan sosial dan politik sering kali memicu penyebaran ujaran kebencian, terutama yang berkaitan dengan isu-isu SARA (Suku, Agama, Ras, dan Antar-golongan). Penyebaran ujaran kebencian yang tidak terkendali dapat menimbulkan konflik sosial yang lebih luas dan berdampak negatif terhadap individu maupun masyarakat secara keseluruhan.

Mendeteksi dan mencegah penyebaran ujaran kebencian di media sosial menjadi tantangan besar, terutama karena besarnya jumlah data yang dihasilkan setiap hari. Metode konvensional yang bergantung pada pelaporan manual dan moderasi oleh manusia sering kali tidak efisien dan tidak mampu mengimbangi volume data yang terus meningkat. Oleh karena itu, diperlukan solusi otomatis yang dapat mengidentifikasi ujaran kebencian dengan lebih cepat dan akurat.

Dalam bidang Machine Learning, algoritma Support Vector Machine (SVM) telah terbukti efektif dalam berbagai tugas klasifikasi teks. SVM bekerja dengan mencari hyperplane optimal yang dapat memisahkan dua kelas data dalam ruang fitur berdimensi tinggi. Keunggulan SVM terletak pada kemampuannya dalam menangani data berdimensi tinggi dan non-linear, serta ketahanannya terhadap overfitting. Keunggulan ini menjadikan SVM sebagai salah satu metode yang layak untuk diterapkan dalam deteksi ujaran kebencian.

Namun, penerapan SVM dalam konteks deteksi ujaran kebencian di Twitter, khususnya dalam bahasa Indonesia, masih menghadapi beberapa tantangan. Bahasa Indonesia memiliki karakteristik unik, seperti variasi dialek, penggunaan bahasa gaul, serta konteks budaya yang spesifik, yang dapat mempersulit pemrosesan bahasa alami (Natural Language Processing/NLP). Selain itu, penelitian sebelumnya lebih banyak berfokus pada penggunaan deep learning dalam mendeteksi ujaran kebencian, sementara eksplorasi terhadap efektivitas SVM dalam konteks bahasa Indonesia masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi efektivitas SVM dalam mendeteksi ujaran kebencian pada tweet berbahasa Indonesia.

Lebih lanjut, penelitian ini juga mempertimbangkan integrasi SVM dengan teknologi terkini, seperti pemrosesan big data dan NLP, untuk meningkatkan akurasi deteksi ujaran kebencian. Beberapa studi sebelumnya telah menunjukkan keberhasilan penggunaan SVM dalam kombinasi dengan teknik NLP untuk klasifikasi teks, namun belum banyak yang mengeksplorasi penerapannya secara spesifik dalam konteks bahasa Indonesia. Selain itu, penelitian ini menggunakan dataset yang dirancang khusus untuk mencerminkan keunikan bahasa dan budaya Indonesia, sehingga model yang dikembangkan dapat lebih relevan dan akurat dalam mendeteksi ujaran kebencian di Twitter.

Dengan demikian, penelitian ini memiliki kontribusi yang signifikan dalam upaya mitigasi penyebaran ujaran kebencian di media sosial. Melalui pengembangan dan evaluasi model SVM yang disesuaikan dengan karakteristik bahasa dan konteks sosial Indonesia, diharapkan penelitian ini dapat memberikan solusi yang lebih efektif dalam menangani masalah ujaran kebencian secara otomatis dan berkelanjutan.

Berdasarkan penelitian Mohammad Attar Jibran, Ade Eviyanti, dan Y. Findawati (2023) mengenai “Hate Speech Detection Using Support Vector Machine (SVM) Method [Deteksi Ujaran Kebencian Menggunakan Metode Support Vector Machine (SVM)]” [1]. Dalam penelitiannya juga menggunakan metode XGBoost dan SVM with Randomized Search Cross Validation sebagai metode banding. Pada penelitian ini menggunakan dataset dari Twitter untuk mendeteksi ujaran kebencian dengan menggunakan metode SVM. Hasil penelitian, yang menggunakan data latih 90% dan data uji 10%, menunjukkan akurasi yang lebih tinggi dibandingkan dengan metode XGBoost dan with Randomized Search Cross Validation (RSCV) sebesar 95,87% dan 87,30% untuk data latih dan uji.

Pada penelitian Kristiawan Nugroho, Endang Tjahjaningsih, Lie Liana, dan Raden Mohammad Herdian Bhakti (2023) mengenai “Prediksi Ujaran Kebencian Berbasis Text Pada Sosial Media Menggunakan Metode Neural Network” [2]. Dataset yang digunakan pada penelitian ini adalah Indonesian abusive and hate speech pada Twitter yang didapat dari kaggle. Hasil penelitian yang didapatkan menggunakan metode Neural Network (NN) mendapatkan hasil skor 73% yang lebih besar dibandingkan metode Decision Tree yang mendapatkan hasil 72.7%, dan metode KNN yang mendapatkan skor 68%.

Pada Penelitian Juan Kalyzta, Muhammad Ardi Willdan, Selfiana Halfiani, Indra (2022), mengenai “PENERAPAN ANALISIS SENTIMEN UJARAN KEBENCIAN TERHADAP VAKSINASI COVID-19 PADA TWEET BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR” [3]. Data dalam penelitian ini bersumber dari media sosial Twitter yang diperoleh melalui proses Crawling menggunakan Tweepy. Data tweet dibagi menjadi data latih sebesar 90% dan data uji sebesar 10%. Pada penelitian ini menggunakan dataset cuitan Twitter yang dikumpulkan menggunakan Twitter API yang berjumlah 100 cuitan. Hasil yang didapatkan setelah melakukan preprocessing text dan penerapan algoritma Random Forest dengan data latih sebesar 70% dan data uji 30% mendapatkan accuracy 35%, precision 20%, dan recall 100% dengan Nilai

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.219-226

K=3.

Pada penelitian Jasman Pardede dan Rangga Alfyansyah (2024), yang berjudul "IDENTIFIKASI UJARAN KEBENCIAN PADA SOSIAL MEDIA BAHASA INGGRIS MENGGUNAKAN RECURSIVE NEURAL NETWORK" [4]. Menggunakan dataset dari Kaggle yang di dapat dari media sosial Twitter yang berjudul "Hate speech offensive tweets". Hasil yang didapat dengan menggunakan RNN mendapatkan nilai precision, recall, accuracy, dan f-measure mencapai 0.78, 0.74, 0.76, dan 0.76.

Support Vector Machine (SVM) adalah algoritma yang dikembangkan oleh Vladimir Vapnik dan rekannya pada awal 1990-an, yang bertujuan untuk mencari hyperplane optimal yang memisahkan dua kelas data dengan margin terbesar [5]. Margin adalah jarak antara hyperplane dan titik data terdekat dari kedua kelas, dan maksimalisasi margin ini penting untuk meningkatkan kemampuan generalisasi model [6]. SVM juga menggunakan support vector, yaitu titik data terdekat yang menentukan posisi hyperplane, serta teknik kernel yang memungkinkan pemisahan data non-linear dengan memetakan data ke ruang dimensi lebih tinggi tanpa perhitungan eksplisit [7]. Hal ini menjadikan SVM efektif untuk klasifikasi teks, di mana setiap dokumen direpresentasikan sebagai vektor fitur [8], dan SVM digunakan untuk memisahkan teks ke dalam kategori seperti ujaran kebencian dan bukan ujaran kebencian [9].

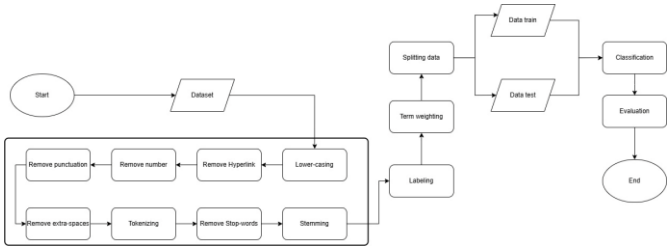
Ujaran kebencian adalah ekspresi yang mengandung kebencian atau diskriminasi terhadap individu atau kelompok tertentu berdasarkan karakteristik seperti ras, agama, dan orientasi seksual, dan diatur dalam Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) di Indonesia [10]. Karakteristik ujaran kebencian di media sosial meliputi bahasa provokatif, target spesifik, penyebaran cepat, anonimitas, dan reaksi berantai [11]. Natural Language Processing (NLP) berperan penting dalam mendeteksi ujaran kebencian dengan membantu ekstraksi fitur dari teks, melalui langkah-langkah seperti tokenisasi dan stemming [12][13][14]. Dalam penelitian ini, klasifikasi teks digunakan untuk mendeteksi ujaran kebencian dalam tweet berbahasa Indonesia dengan menerapkan SVM, diharapkan dapat mencapai akurasi tinggi dalam identifikasi konten berbahaya [15][16].

II. METODE PENELITIAN

Metode Penelitian ini bertujuan untuk menjelaskan langkah-langkah dan pendekatan yang digunakan dalam penelitian ini untuk menerapakan algoritma support vector machine dalam mendeteksi ujaran kebencian dalam media social twitter. Diharapkan penjelasan yang diberikan tentang tahapan penilitan, pengumpulan data, serta penerapan teori dan algoritma yang dipilih dengan tujuan menguji efektivitas support vector machine dalam mendeteksi ujaran kebencian

A. Tahapan Penelitian

Berikut tahapan penelitian dalam penerapan algoritma support vector machine dalam mendeteksi ujaran



Gambar. 1. Tahapan Penelitian

kebencian dalam media social twitter.

B. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini terdiri dari tweet yang diambil dari platform media sosial Twitter, dengan fokus pada deteksi ujaran kebencian. Dataset ini berisi total 1564 baris yang merupakan tweet yang di kumpulkan dari tanggal 1 Januari 2024 hingga 19 November 2024. Contoh tweet yang ada di dataset ini mencakup berbagai topik, seperti pernyataan politik, komentar tentang politik, pilpres, dan timnas Indonesia. Dengan menggunakan dataset ini, penelitian bertujuan untuk menganalisis dan mengembangkan model klasifikasi yang efektif untuk mendeteksi ujaran kebencian dalam konteks media sosial, sehingga dapat memberikan kontribusi terhadap pemahaman dan penanganan isu tersebut.

C. Labeling Data

Setiap tweet yang diambil kemudian diberi label dengan kata kunci ujaran kebencian menggunakan fungsi classify hate speech, yang menganalisis teks berdasarkan daftar kata kunci yang telah ditentukan sebelumnya. Fungsi

ini memeriksa apakah teks mengandung salah satu dari kata-kata yang termasuk dalam daftar kata kunci ujaran kebencuan. Jika ditemukan kata yang sesuai, maka tweet tersebut akan diklasifikasikan sebagai 'hate speech'. Sebaliknya, jika tidak ada kata yang cocok, tweet akan diberi label 'no hate speech'. Proses ini memungkinkan identifikasi dan pemisahan tweet yang mengandung ujaran kebencian dari yang tidak.

D. Term Weighting

Term weighting, atau tahapan untuk memberikan bobot pada tiap kata dalam masing-masing baris data, atau dalam hal ini dokumen tweet. Pada tahap ini, jenis term weighting yang digunakan adalah dengan perhitungan TF-IDF (Term Frequency-Inverse Document Frequency). Di bawah ini adalah rumus yang digunakan:

$$TF\ (t, d) = jumlah\ kemunculan\ t\ dalam\ d\ jumlah\ kata\ dalam\ d$$
$$IDF\ (t, D) = \log\ log\ jumlah\ total\ dokumen\ (jumlah\ dokumen\ yang\ mengandung\ t) \quad (1)$$
$$TF - IDF\ (t, d, D) = TF\ (t, d) \times IDF\ (t, D)\ [17]$$

Perhitungan Term Frequency dengan rumus di atas berguna untuk menghitung seberapa pentingnya masing-masing kata (atau dalam rumus disimbolkan dengan t) dalam satu dokumen, atau dalam hal ini satu tweet (baris data dalam dataset atau dalam rumus disimbolkan dengan d), dimana semakin tinggi nilai TF, maka semakin besar juga bobot kata itu dalam satu tweet itu. Di sisi lain, perhitungan Inverse Document Frequency dilakukan guna melihat seberapa sering suatu kata muncul dalam keseluruhan koleksi dokumen (keseluruhan dataset atau dalam rumus disimbolkan dengan D), dimana semakin sering kata itu muncul, semakin rendah nilai IDF, semakin kurang penting bobot kata itu dalam dokumen.

E. Splitting Data

Pada tahap splitting data dataset yang diperoleh dari tweet Twitter akan dibagi menjadi dua bagian, yaitu data latih (training) dan data uji (testing). Tahap ini penting untuk memastikan algoritma Support Vector Machine (SVM) yang dikembangkan dapat berfungsi dengan baik terhadap data baru yang belum pernah dilihat sebelumnya. Dalam penelitian ini, pembagian data yang digunakan adalah 75:25, di mana 75% dari total dataset akan digunakan sebagai data training dan 25% data digunakan sebagai data test. Pembagian ini dilakukan menggunakan library sklearn, yang menghasilkan variabel Train_X dan Test_X untuk data tweet, serta Train_Y dan Test_Y untuk label. Dengan cara ini, diharapkan model SVM dapat memberikan klasifikasi yang akurat dalam mendeteksi ujaran kebencian di media sosial Twitter.

F. Classification

Classification, atau proses klasifikasi adalah suatu metode dalam Machine Learning yang digunakan untuk mengkategorikan data ke dalam label tertentu berdasarkan juga fitur-fitur tertentu guna melatih suatu model untuk dapat mempelajari pola antar fitur. Model yang telah dilatih ini kemudian diharapkan dapat melakukan prediksi terhadap daru yang belum dilabel. Berdasarkan data yang dilatih, ada dua jenis klasifikasi, yaitu klasifikasi dengan data yang sudah dilabeli (Supervised Learning), dan klasifikasi dengan data yang belum dilabeli (Unsupervised Learning). Dalam penelitian ini, metode klasifikasi yang akan digunakan adalah Supervised Learning, dengan algoritma Support Vector Machine [18].

Support Vector Machine (SVM) bekerja dengan menemukan hyperplane optimal yang memisahkan data ke dalam kelas-kelas yang berbeda. SVM beroperasi dengan memetakan data ke dalam ruang berdimensi tinggi dan kemudian mencari hyperplane yang memaksimalkan margin, yaitu jarak terdekat antara titik data dari kedua kelas. Titik-titik data yang berada di batas margin disebut support vectors. SVM dapat menggunakan kernel trik untuk mengubah data non-linear ke ruang berdimensi lebih tinggi, sehingga memungkinkan pemisahan yang lebih baik. Dengan demikian, SVM efektif dalam menghasilkan klasifikasi yang akurat, terutama dalam kasus di mana data tidak dapat dipisahkan secara linear [19].

G. Evaluation

Setelah model dilatih dengan data train, kami melakukan evaluasi menggunakan data test untuk mengukur seberapa baik model dapat mengklasifikasikan tweets sebagai ujaran kebencian atau bukan. Dalam tahap ini, akan meilhat kinerja metode klasifikasi Support Vector Machine dengan menganalisa mulai dari confusion matrix serta dengan menggunakan beberapa metrik evaluasi, termasuk akurasi, recall, presisi, dan F1 Score.Pertama, confusion matrix akan menunjukkan jumlah prediksi yang benar dan salah yang dibuat oleh model, terdiri dari empat komponen: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Ini memberikan gambaran rinci tentang kesalahan yang dibuat oleh model. Selanjutnya, akurasi, recall, presisi, serta F1 score akan dihitung dengan rumus:

$$Akurasi = TP+TN\ TP+TN+FP+FN \times 100\%$$
$$Recall = TP\ TP+FN \times 100\% \quad (1)$$
$$Presisi = TP\ TP+FP \times 100\%$$
$$F1\ Score = 2 \times Presisi \times Recall\ Presisi+Recall \times 100\%$$ [20]

Model SVM dengan nilai akurasi, recall, precision, dan F1 score yang tinggi dianggap memiliki kinerja yang lebih baik dalam mengklasifikasikan data dengan benar.

III. HASIL PEMBAHASAN

A. PENGUMPULAN DATA

Proses pengumpulan data dilakukan dengan mengambil tweet-tweet yang mengandung ujaran kebencian dari platform Twitter menggunakan API Twitter. Dataset yang digunakan dalam penelitian ini terdiri dari 1.567 tweet yang diambil secara acak dengan pencarian kata kunci terkait ujaran kebencian, seperti "politik", "pilpres", dan "politik Indonesia". Penggunaan kata kunci ini bertujuan untuk memastikan bahwa data yang dikumpulkan relevan dengan konteks ujaran kebencian yang sering muncul dalam diskusi politik di Indonesia.

B. PREPROCESSING DATA

Proses pra-pemrosesan data terdiri dari beberapa langkah, termasuk lower casing, penghapusan hyperlink, penghapusan angka, penghapusan tanda baca, penghapusan spasi ekstra, pemisahan teks menjadi kata-kata (tokenizing), penghapusan stopwords, dan penghapusan awalan dan akhiran kata (stemming). Setiap langkah pra-pemrosesan memiliki dampak signifikan terhadap kinerja model. Misalnya, penghapusan stopwords membantu mengurangi noise dalam data, sehingga model dapat lebih fokus pada kata-kata yang lebih informatif. Stemming juga berkontribusi pada peningkatan akurasi klasifikasi dengan mengurangi variasi kata yang memiliki makna serupa. Penelitian ini menunjukkan bahwa langkah-langkah tertentu, seperti penghapusan stopwords dan stemming, memberikan kontribusi signifikan terhadap peningkatan performa model, karena memungkinkan model untuk lebih baik dalam mengenali pola yang relevan dalam data. Contoh data hasil pra-pemrosesan yang dapat dilihat pada Tabel 1.

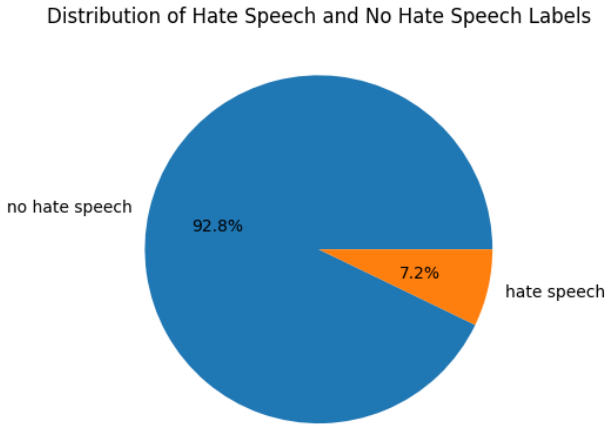
TABEL I
HASIL PREPROCESSING DATA

No	Data Tweets
1	['bila nepotisme tidak boleh mengapa presiden cara resmi justru manfaat dalam pemilu presiden']
2	['panggil main timnas putra senior indonesia afc bahrain tiongkok']
3	['indonesia belum siap untuk demokrasi kata prabowo kepada saya jurnalis allan nairn dalam temu itu negara ini kata butuh rezim otoriter yang jinak']
4	['timnas indonesia telah tiba di china terima kasih temanteman ajar atau mahasiswa indonesia di china yang sambut datang skuat garuda']
5	['timnas indonesia telah tiba di china terima kasih temanteman ajar atau mahasiswa indonesia di china yang sambut datang skuat garuda']

C. LABELING DATA

Labeling data adalah proses memberi label pada data untuk memberikan konteks dan kategori. Dapat dilihat pada Gambar 3 akan disajikan distribusi label data. Proses ini penting untuk memastikan bahwa model dapat belajar dari data yang terstruktur dengan baik, sehingga dapat meningkatkan akurasi klasifikasi.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.219-226



Gambar. 2. Distribusi Labeling Data

D. *TERM WEIGHTING*

Dapat dilihat pada Gambar 3, hasil TF-IDF membantu mengidentifikasi kata-kata kunci yang signifikan dalam setiap dokumen. Kata-kata dengan nilai TF-IDF tinggi, seperti "fuck" dan "aku", menunjukkan bahwa kata-kata ini sering muncul dalam konteks ujaran kebencian. Kata "fuck", misalnya, dapat menunjukkan emosi yang kuat dan sering digunakan dalam konteks negatif, sedangkan "aku" mungkin menunjukkan subjek yang terlibat dalam pernyataan kebencian. Pemahaman tentang kata-kata ini penting karena dapat memengaruhi kemampuan model untuk mengklasifikasikan tweet dengan akurat. Kata-kata dengan nilai TF-IDF rendah atau nol, seperti "jendela" atau "gusar", menunjukkan bahwa kata-kata ini tidak memberikan banyak informasi dalam konteks dokumen secara umum, sehingga tidak berkontribusi pada klasifikasi.

E. *SPLITTING DATA*

Data secara keseluruhan kemudian dipisahkan untuk dilatih serta diuji untuk model klasifikasi yang akan digunakan, yaitu Support Vector Machine (SVM). Model ini memiliki pembagian data, yaitu 75% untuk training

	itu	ga	['aku	tapi	off	fuck	adab	\
0	0.109882	0.093354	0.100685	0.042885	0.107197	0.107197	0.118336	
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
...	...	...	...	...	...	...	...	
1561	0.049447	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
1562	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
1563	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
1564	0.024121	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
1565	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

	baik	nyerang	terlalu	...	jagain']	waras	['youtuber	\
0	0.054423	0.118336	0.081391	...	0.000000	0.000000	0.000000	
1	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
2	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
4	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
...	...	...	...	...	...	...	...	
1561	0.000000	0.000000	0.000000	...	0.079877	0.000000	0.000000	
1562	0.000000	0.000000	0.000000	...	0.000000	0.088752	0.000000	
1563	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.355008	
1564	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
1565	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	

	jendela	gusar	sdg	inisiasi	treshhold	['yuk	minzu
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
...	...	...	...	...	...	...	...
1561	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1562	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1563	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1564	0.077929	0.077929	0.077929	0.077929	0.077929	0.000000	0.000000
1565	0.000000	0.000000	0.000000	0.000000	0.000000	0.228219	0.228219

[1566 rows x 7095 columns]

Gambar. 3. Hasil TF-IDF

dan 25% untuk testing. Pembagian ini dirancang untuk memastikan bahwa model memiliki cukup data untuk training dan dapat diuji secara efektif pada data yang sudah di labelling.

F. CLASSIFICATION

Hasil ini dianalisis menggunakan metrik evaluasi utama, yaitu akurasi dan recall. Temuan dari klasifikasi ini memiliki implikasi praktis yang sangat penting. Misalnya, platform media sosial dan pemerintah dapat memanfaatkan hasil ini untuk lebih efektif dalam mengidentifikasi dan mengurangi penyebaran ujaran kebencian di dunia maya.

TABEL 2
HASIL KLASIFIKASI

Metrik Evaluasi	Hasil
Accuracy	97.19%
Recall	97.19%
Precision	97.28%
F1 Score	96.82%

Hasil penelitian ini membuka peluang untuk pengembangan model deteksi ujaran kebencian yang lebih baik di masa depan. Salah satu cara untuk meningkatkan akurasi adalah dengan menerapkan teknik augmentasi data, seperti menggunakan sinonim atau variasi dalam struktur kalimat, yang dapat memperkaya dataset. Selain itu, menjelajahi arsitektur model yang lebih kompleks, seperti CNN atau RNN, serta memanfaatkan model pre-trained seperti BERT, dapat membantu model lebih efektif dalam memahami konteks. Peningkatan dalam proses pra-pemrosesan dan penambahan fitur relevan, seperti analisis sentimen, juga dapat memberikan wawasan tambahan yang berguna. Melakukan evaluasi yang lebih mendalam melalui validasi silang dan mengumpulkan lebih banyak data dari berbagai konteks akan membantu model belajar dari beragam variasi ujaran kebencian. Dengan menerapkan strategi-strategi ini, diharapkan model dapat lebih efektif dalam mengidentifikasi dan mengurangi penyebaran ujaran kebencian di platform media sosial.

Terakhir, perlu dibahas kelebihan dan kelemahan SVM dalam konteks penelitian ini. SVM dipilih sebagai algoritma utama karena kemampuannya dalam menangani data berdimensi tinggi dan non-linear, serta ketahanannya terhadap overfitting. Namun, ada tantangan khusus dalam menerapkan SVM untuk deteksi ujaran kebencian dalam bahasa Indonesia, seperti variasi dialek dan penggunaan bahasa gaul yang dapat memengaruhi akurasi klasifikasi. Penelitian ini diharapkan dapat memberikan wawasan lebih dalam tentang bagaimana SVM dapat diadaptasi untuk konteks lokal dan tantangan yang dihadapi dalam mendeteksi ujaran kebencian.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Support Vector Machine (SVM) dalam mendeteksi ujaran kebencian pada tweet berbahasa Indonesia. Dataset yang digunakan terdiri dari 1.564 tweet yang dikumpulkan dari media sosial Twitter. Hasil penelitian menunjukkan bahwa algoritma SVM mampu mencapai akurasi tinggi dalam mendeteksi ujaran kebencian, dengan tingkat akurasi sebesar 97,19%. Pada hasil ini menunjukan bahwa algoritma SVM dapat menjadi alat yang efektif untuk mengidentifikasi ujaran kebencian dalam tweet berbahasa Indonesia. Selain itu, penelitian ini juga membuktikan bahwa teknik pra-pemrosesan data yang tepat dapat meningkatkan akurasi algoritma dalam mendeteksi ujaran kebencian.

V. DAFTAR PUSTAKA

[1] M. A. Jibran, A. Eviyanti, and Y. Findawati, "Deteksi Ujaran Kebencian Menggunakan Metode Support Vector Machine (SVM)," *Kesatria : Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 4, no. 4, 2023.

[2] K. Nugroho, E. Tjahjaningsih, L. Liana, and R. Mohamad Herdian Bhakti, "Prediksi Ujaran Kebencian Berbasis Text Pada Sosial Media Menggunakan Metode Neural Network," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 5, no. 1, 2023, doi: 10.46772/intech.v5i1.1063.

[3] J. Kalyzta, M. A. Willdan, S. Halfiani, and I. Indra, "PENERAPAN ANALISIS SENTIMEN UJARAN KEBENCIAN TERHADAP VAKSINASI COVID-19 PADA TWEET BERBAHASA INDONESIA MENGGUNAKAN ALGORITME K-NEAREST NEIGHBOR," *IDEALIS : InDonEsiA journaL Information System*, vol. 5, no. 2, pp. 87–97, Jul. 2022, doi: 10.36080/idealis.v5i2.2959.

[4] J. Pardede and R. Alfyansyah, "Identifikasi Ujaran Kebencian Pada Sosial Media Bahasa Inggris Menggunakan Recursive Neural Network," *Komputa : Jurnal Ilmiah Komputer dan Informatika*, vol. 13, no. 1, pp. 23–32, Apr. 2024, doi: 10.34010/komputa.v13i1.10783.

[5] SINAGA, Novendra Adisaputra; HAYADI, B Herawan; SITUMORANG, Zakarias. PERBANDINGAN AKURASI ALGORITMA NAÏVE BAYES, K-NN DAN SVM DALAM MEMPREDIKSI PENERIMAAN PEGAWAI. Jurnal Tekinkom (Teknik Informasi dan Komputer), [S.l.], v. 5, n. 1, p. 27-34, june 2022. ISSN 2621-3079.

[6] Pernama, B., & Purnomo, H. D. Analisis Risiko Pinjaman dengan Metode Support Vector Machine, Artificial Neural Network dan Naïve Bayes. Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi), 7(1), 92–99. https://doi.org/10.35870/jtik.v7i1.693

[7] Fadillah, V., Hamami, F., & Andreswari, R. "Analisis Sentimen Berbasis Aspek Terhadap Ulasan Pengguna Aplikasi Pegadaian Digital Dengan Multiclass Multioutput Menggunakan Algoritma Support Vector Machine." KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen), vol. 4, no. 4, pp. 977-987. ISSN: 2720-992X.

[8] Ahmad, A., & Gata, W. Sentimen Analisis Masyarakat Indonesia di Twitter Terkait Metaverse dengan Algoritma Support Vector Machine. Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi), 6(4), 548–555. https://doi.org/10.35870/jtik.v6i4.569

[9] Dewi, R. A., & Santoso, H. "Deteksi Ujaran Kebencian Menggunakan Algoritma Naïve Bayes dan Support Vector Machine," Jurnal Ilmiah Teknologi Informasi, vol. 10, no. 1, pp. 45–52, Jan. 2022.

[10] F. Abdusyukur, "PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI PENCEMARAN NAMA BAIK DI MEDIA SOSIAL TWITTER," Komputa : Jurnal Ilmiah Komputer dan Informatika, vol. 12, no. 1, pp. 73–82, May 2023, doi: 10.34010/komputa.v12i1.9418.

[11] K. A. Kocoń, A. Figas, M. Gruza, "Offensive, Aggressive, and Hate Speech Analysis: From Data-Centric to Human-Centered Approach," Information Processing & Management, vol. 58, no. 5, pp. 102643, 2021.

[12] A. K. Sharma, S. K. Gupta, "Natural Language Processing Techniques for Text Classification: A Review," International Journal of Computer Applications, vol. 975, no. 8887, pp. 1-6, 2021.

[13] Kusuma, A. P., & Setiawan, A. "Implementasi Metode Machine Learning untuk Deteksi Ujaran Kebencian di Twitter," Jurnal Informatika, vol. 18, no. 2, pp. 89–96, Apr. 2023.

[14] Halim, A., & Rahmawati, N. "Penggunaan Algoritma Klasifikasi untuk Deteksi Ujaran Kebencian dalam Teks Berbahasa Indonesia," Jurnal Teknologi Informasi dan Komunikasi, vol. 11, no. 1, pp. 34–42, Feb. 2022.

[15] Buana, R. S., Gata, W., Widodo, A. Z. P., Setiawan , H. ., & Hilyati, K. (2023). Analisis Sentimen pada Komen Twitter Pawang Hujan Mandalika dengan Support Vector Machine (SVM) dan Naïve Bayes. Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi), 7(2), 194–200. https://doi.org/10.35870/jtik.v7i2.705

[16] Sari, D. P., & Hidayat, R. "Penerapan Metode Support Vector Machine untuk Klasifikasi Ujaran Kebencian di Media Sosial," Jurnal Teknologi dan Sistem Komputer, vol. 9, no. 2, pp. 123–130, Jun. 2021.

[17] Ibnu Surya Wibowo, Arita Winanti, and Indah Susilawati, "Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF ," Jurnal Teknik Informatika dan Sistem Informasi, vol. 11, pp. 99–111, Mar. 2023.

[18] R. A. F. Alamsyah, D. A. Putra, and S. H. Sari, "Klasifikasi Ujaran Kebencian Menggunakan Metode Support Vector Machine (SVM) dan Naïve Bayes," Jurnal Teknologi dan Sistem Komputer, vol. 9, no. 2, pp. 123-130, 2022.

[19] M. Fadli and R. A. Saputra, "KLASIFIKASI DAN EVALUASI PERFORMA MODEL RANDOM FOREST UNTUK PREDIKSI STROKE," *JT: Jurnal Teknik*, vol. 12, no. 02, 2023.

[20] Fu rqan, M., Ikhsan, M., & Aini, R.. "Algoritma Support Vector Machine Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Pandemi Virus Corona Di Media Sosial." KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen), vol. 4, no. 4, pp. 908-915. doi: 10.30645/kesatria.v4i4.241.