

Analisis Sentimen Masyarakat Indonesia Terkait Konten Deepfake Pada Platform X Menggunakan Algoritma SVM

Muhammad Reza Maulana ^{1*}, Agung Nugroho ², Asep Suprianto ³

¹ Universitas Pelita Bangsa; rezamaulana9569@gmail.com

² Universitas Pelita Bangsa; agung@pelitabangsa.ac.id

³ Universitas Pelita Bangsa; asepsupriyanto@pelitabangsa.ac.id

Abstrak: Maraknya penyebaran konten deepfake di platform media sosial memunculkan kekhawatiran tentang kemungkinan adanya disinformasi dan manipulasi informasi di ranah digital. Platform media sosial X menjadi salah satu lokasi utama bagi masyarakat untuk menyampaikan pandangan mereka tentang fenomena ini. Penelitian ini bertujuan untuk menganalisis pandangan masyarakat Indonesia mengenai deepfake di platform X dan menilai kinerja algoritma Support Vector Machine (SVM) dalam mengklasifikasikan sentimen. Data diperoleh dengan teknik crawling menggunakan kata kunci “deepfake” pada periode Januari 2024 hingga Juni 2025, menghasilkan 2.116 tweet dan 1.841 data bersih setelah dilakukan preprocessing. Pelabelan data dilakukan dengan pendekatan lexicon-based, dengan ekstraksi fitur menggunakan TF-IDF. Model klasifikasi menggunakan SVM dengan kernel linear serta pembagian data training dan testing dengan rasio 80:20. Evaluasi model menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa pendapat masyarakat lebih banyak didominasi oleh sentimen negatif. Model SVM mencatat akurasi 80,76%, presisi 68,81%, recall 66,96%, dan F1-score 67,87%, yang menunjukkan performa klasifikasi pada tingkat sedang dalam skema pelabelan berbasis leksikon. Hasil ini dapat digunakan sebagai dasar untuk memantau opini publik terkait fenomena deepfake serta mendukung pengembangan strategi literasi digital di media sosial.

Keywords: Analisis Sentimen; Deepfake; Support Vector Machine; TF-IDF; Lexicon-Based

DOI: <https://doi.org/10.47134/jacis.v6i1.180>

*Correspondensi: Muhammad Reza Maulana

Email: rezamaulana9569@gmail.com

Receive: 15 Maret 2026

Accepted: 13 April 2026

Published: 22 April 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstrak: The widespread circulation of deep fake content on social media platforms has raised concerns about the potential for disinformation and information manipulation in the digital realm. The social media platform X has become one of the primary venues for the public to express their views on this phenomenon. This study aims to analyze the views of the Indonesian public regarding deepfakes on the X platform and evaluate the performance of the Support Vector Machine (SVM) algorithm in classifying sentiment. Data was collected using a crawling technique with the keyword “deepfake” from January 2024 to June 2025, yielding 2,116 tweets and 1,841 clean data points after preprocessing. Data labeling was performed using a lexicon-based approach, with feature extraction using TF-IDF. The classification model utilized an SVM with a linear kernel, and the training and testing data were split in an 80:20 ratio. Model evaluation employed a confusion matrix with metrics including accuracy, precision, recall, and F1-score. The results

indicate that public opinions are predominantly dominated by negative sentiment. The SVM model recorded an accuracy of 80.76%, precision of 68.81%, recall of 66.96%, and an F1-score of 67.87%, indicating moderate classification performance within the lexicon-based labeling framework. These results can serve as a foundation for monitoring public opinion regarding the deepfake phenomenon and supporting the development of digital literacy strategies on social media.

Keywords: Sentiment Analysis; Deepfake; Support Vector Machine; TF-IDF; Lexicon-Based

PENDAHULUAN

Perkembangan teknologi digital dalam beberapa tahun terakhir telah mengalami lonjakan yang luar biasa, terutama di bidang kecerdasan buatan (*Artificial Intelligence*). Salah satu inovasi yang kini menarik perhatian dunia adalah teknologi deepfake, yang merupakan teknik untuk memanipulasi video atau audio menggunakan algoritma AI guna mengubah wajah atau suara seseorang sehingga tampak seolah-olah nyata. Dalam beberapa tahun terakhir, penyalahgunaan deepfake semakin meningkat dan telah dipakai dalam berbagai situasi, termasuk penyebaran berita palsu, penipuan secara online, serta manipulasi konten politik di platform media sosial, yang menyebabkan turunnya kepercayaan masyarakat terhadap informasi digital [1][2]. Awalnya, teknologi ini dimanfaatkan dalam industri hiburan dan produksi kreatif, tetapi seiring waktu, deepfake juga menimbulkan kekhawatiran karena dapat digunakan untuk menyebarkan disinformasi, penipuan digital, serta merusak reputasi. Situasi ini menunjukkan bahwa kemajuan AI tidak hanya membawa keuntungan, tetapi juga dapat menjadi risiko bagi keaslian informasi digital [3].

Di Indonesia, penggunaan media sosial menunjukkan perkembangan yang sangat signifikan. Menurut laporan *We Are Social 2024*, jumlah pengguna aktif platform X (*Twitter*) di negara Indonesia mencapai 24,69 juta, menjadikannya salah satu media sosial yang banyak digunakan untuk menyebarkan informasi dan berdiskusi di publik [4]. Ciri khas platform X yang terbuka dan berbasis opini membuatnya menjadi sumber yang relevan untuk menganalisis pandangan masyarakat tentang isu teknologi, termasuk fenomena deepfake. Rendahnya tingkat literasi digital di kalangan masyarakat dapat meningkatkan risiko penyebaran konten deepfake, karena orang-orang sering mengalami kesulitan dalam membedakan antara konten yang asli dan yang telah dimanipulasi secara digital, sehingga dapat mempercepat penyebaran disinformasi di dunia digital [5][6]. Oleh karena itu, pendekatan berbasis data sangat diperlukan untuk memahami bagaimana masyarakat Indonesia merespons dan memandang keberadaan teknologi deepfake.

Analisis sentimen adalah salah satu teknik yang sering dipakai untuk mengidentifikasi pandangan masyarakat terkait isu tertentu melalui data teks di platform media sosial. Teknik ini memfasilitasi pengelompokan pandangan masyarakat ke dalam kategori sentimen seperti positif maupun negatif secara otomatis. Dalam perkembangan terbaru, metode berbasis *deep learning* seperti *Bidirectional Encoder Representations from Transformers* (BERT) telah memperlihatkan kinerja yang sangat baik dalam penelitian analisis sentimen karena kemampuannya untuk memahami konteks bahasa secara mendalam [7]. Namun, metode tersebut umumnya membutuhkan sumber daya komputer yang besar dan dataset yang lebih kompleks. Maka dari itu, Algoritma *Support Vector Machine* (SVM) masih menjadi pilihan

yang relevan karena mampu mengelola data berdimensi tinggi seperti teks dan memiliki kompleksitas yang lebih rendah dengan kinerja yang konsisten [8].

Meskipun ada beberapa penelitian telah dilakukan mengenai analisis sentimen dengan algoritma SVM, sebagian besar fokus pada isu-isu seperti politik, layanan transportasi online, atau kebijakan publik [9]. Penelitian mengenai deepfake umumnya lebih fokus pada deteksi konten dengan menggunakan metode deep learning, serta belum banyak yang melakukan studi analisis sentimen berbasis teks di media sosial, terutama dalam konteks bahasa Indonesia [10]. Selain itu, integrasi metode pelabelan berbasis leksikon dengan algoritma *Machine Learning* seperti SVM masih relatif terbatas dan belum banyak dieksplorasi, terutama dalam konteks analisis deepfake [11][12]. Hal ini menunjukkan adanya kesenjangan penelitian, baik dari segi konteks data, pendekatan metodologi, maupun fokus yang diteliti.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap fenomena deepfake di platform X dengan memanfaatkan algoritma *Support Vector Machine* (SVM). Penelitian ini berkontribusi melalui penyusunan dataset tweet dalam bahasa Indonesia yang berhubungan dengan deepfake, penggunaan metode pelabelan berbasis leksikon dalam analisis sentimen, serta evaluasi kinerja algoritma SVM dengan menerapkan representasi fitur TF-IDF. Hasil dari penelitian ini diharapkan mampu memberikan gambaran empiris tentang pandangan masyarakat terkait deepfake serta berkontribusi dalam pengembangan studi analisis sentimen serta penggunaannya dalam pengawasan opini publik di platform media sosial.

METODE

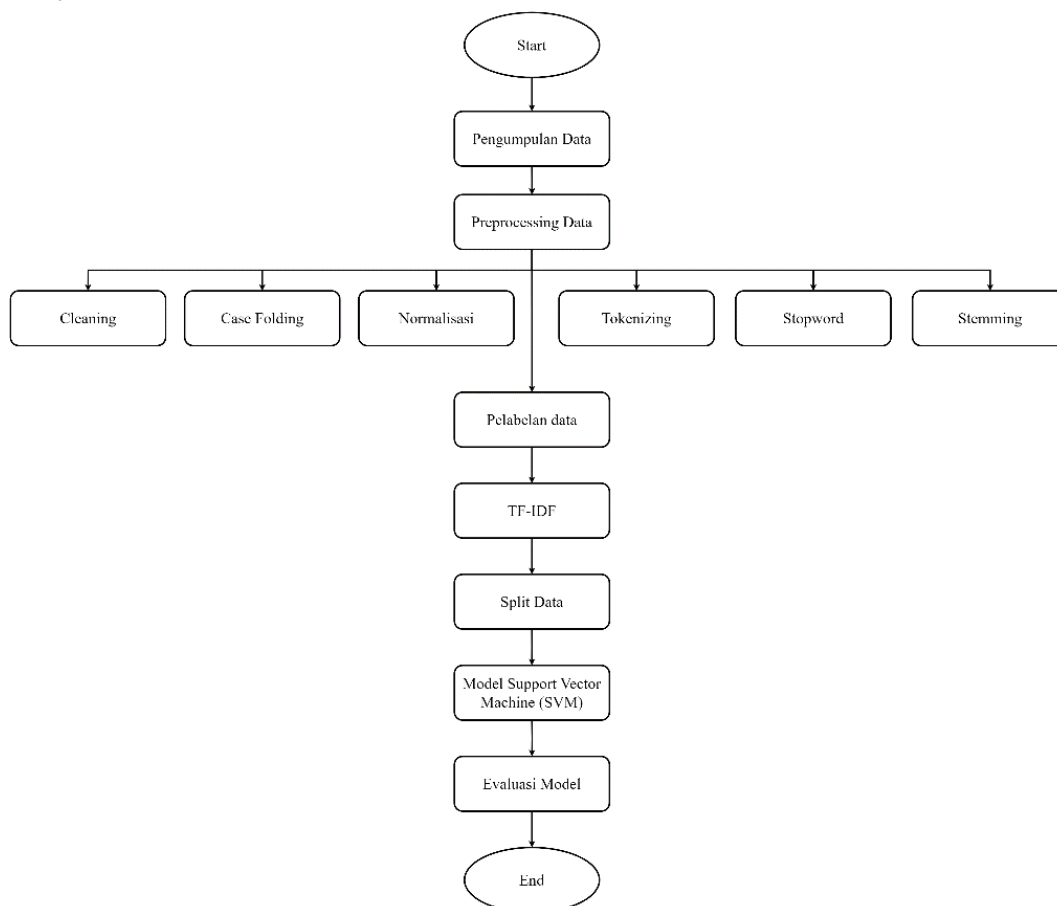
Desain Penelitian

Penelitian ini menerapkan pendekatan kuantitatif dengan metode *text mining* untuk menganalisis sentimen masyarakat Indonesia mengenai fenomena deepfake di platform X. Proses penelitian ini dilakukan melalui langkah-langkah terstruktur, yang mencakup pengumpulan data dengan teknik *crawling*, *preprocessing* teks, pelabelan sentimen menggunakan pendekatan *lexicon-based*, ekstraksi fitur memakai TF-IDF, pembagian dataset menjadi data training dan data testing, serta klasifikasi dengan penggunaan algoritma *Support Vector Machine* (SVM). Dalam penelitian ini, analisis sentimen digunakan untuk mengidentifikasi polaritas opini masyarakat terhadap konten deepfake berdasarkan pada data teks yang diambil dari sosial media. Serangkaian tahapan metode penelitian secara keseluruhan disajikan dalam format diagram pada gambar 1.

1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui teknik *crawling* data di platform media sosial X dengan menggunakan kata kunci “deepfake” dari tanggal 1 Januari 2024 sampai 30 Juni 2025 untuk mendapatkan tweet berbahasa Indonesia yang berkaitan dengan topik penelitian. Proses *crawling* data dilakukan secara otomatis dengan memanfaatkan library pemrograman Python, sehingga data yang diperoleh dapat diproses secara sistematis dan efisien. Data yang diperoleh dapat digunakan untuk analisis, penelitian, atau pengembangan sistem informasi [13][14]. Dalam analisis sentimen, *crawling* data digunakan untuk mengumpulkan tweet atau

komentar mengenai produk atau peristiwa untuk memahami pendapat atau perspektif masyarakat [15].



Gambar 1. Alur metode penelitian

2. Preprocessing Data

Preprocessing dilakukan untuk mempersiapkan data dalam format yang tepat agar bisa dipakai oleh sistem atau algoritma, sehingga proses *knowledge extraction* dapat dilaksanakan [16]. Tahapan *preprocessing* dalam penelitian ini dilakukan secara bertahap yang mencakup *cleaning*, *case folding*, *tokenizing*, *stopword removal*, *stemming*, dan normalisasi [17]. Tahap *cleaning* dilakukan dengan menghilangkan karakter yang tidak penting seperti tanda baca dan simbol. Setelah itu, *case folding* diterapkan untuk mengkonversi semua huruf menjadi huruf kecil. *Tokenizing* dilakukan untuk membagi teks menjadi bagian-bagian kata, setelah itu diikuti dengan penghapusan *stopword* untuk menyingkirkan kata-kata yang tidak memiliki makna signifikan. Proses *stemming* digunakan untuk mengubah kata-kata ke bentuk dasarnya, sementara normalisasi dilakukan untuk mengonversi kata-kata yang tidak baku menjadi kata-kata yang baku [18].

3. Pelabelan Data

Dalam penelitian ini, pelabelan data dilakukan dengan menggunakan pendekatan *lexicon-based* yaitu metode yang mengandalkan sebuah kamus sentimen yang memuat kata-kata positif dan kata-kata negatif, kemudian dipadankan dengan kata-kata dalam kalimat untuk menentukan tingkat polaritasnya [19]. Kamus sentimen yang digunakan

mengacu pada InSet (*Indonesian Sentiment Lexicon*) yang dapat diakses secara publik melalui repositori GitHub <https://github.com/fajri91/InSet.git>.

4. TF-IDF

TF-IDF (*Term Frequency–Inverse Document Frequency*) merupakan metode untuk memberikan bobot pada kata yang berfungsi untuk mengkonversi teks menjadi bentuk numerik dalam vektor fitur. Metode ini menilai tingkat kepentingan sebuah kata dalam suatu dokumen terhadap seluruh kumpulan dokumen atau corpus. TF-IDF terdiri dari dua elemen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF), yang dapat dijelaskan dengan rumus berikut [20]:

$$TF(t, d) = \frac{n_{t,d}}{\text{(Jumlah seluruh kata dalam dokumen)}} \quad (1)$$

$$IDF(t) = \log \frac{N}{df(t)} \quad (2)$$

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

Dimana $n_{t,d}$ adalah frekuensi kemunculan kata t dalam dokumen d ,

N adalah jumlah total dokumen, dan

$df(t)$ adalah jumlah dokumen yang mengandung kata t .

5. Split Data

Pembagian data bertujuan untuk memisahkan dataset ke dalam data *training* dan data *testing* untuk mengevaluasi performa model. Dalam penelitian ini, data tersebut dibagi dengan rasio 80:20, di mana 80% digunakan sebagai data *training* dan 20% sebagai data *testing*. Pembagian data dilakukan secara acak untuk menjaga distribusi data tetap seimbang dan mengurangi masalah overfitting [21].

6. Model Support Vector Machine

Support Vector Machine (SVM) adalah metode klasifikasi yang didasarkan pada supervised learning yang digunakan untuk menentukan kategori berdasarkan pola yang diperoleh dari data training dengan menciptakan garis pemisah (hyperplane) antara kategori yang berbeda [22]. Secara matematis, fungsi keputusan SVM dirumuskan sebagai berikut:

$$f(x) = \text{sign}(w \cdot x + b) \quad (4)$$

Dimana w adalah vektor bobot, x adalah fitur input, dan b adalah bias. Dalam penelitian ini, model SVM menggunakan kernel linear serta parameter `class_weight = balanced` untuk mengatasi masalah ketidakseimbangan data. Proses pengoptimalan parameter dilakukan menggunakan GridSearchCV dengan menerapkan teknik 5-fold cross validation pada parameter C dengan nilai 0.1, 1, dan 10. Algoritma SVM dipilih karena kemampuannya yang baik dalam mengelola data dengan dimensi tinggi seperti teks serta memberikan hasil klasifikasi yang stabil [23].

7. Evaluasi Model

Evaluasi Model dalam penelitian ini menggunakan confusion matrix. Metode ini diterapkan untuk menilai kesesuaian antara hasil prediksi model dan data yang

sebenarnya agar dapat mengevaluasi kemampuan model klasifikasi [24]. Confusion Matrix terdiri dari empat elemen, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) [25]. Berdasarkan elemen-elemen tersebut, kinerja model kemudian diukur dengan menggunakan metrik accuracy, precision, recall, dan F1-score [26].

HASIL DAN PEMBAHASAN

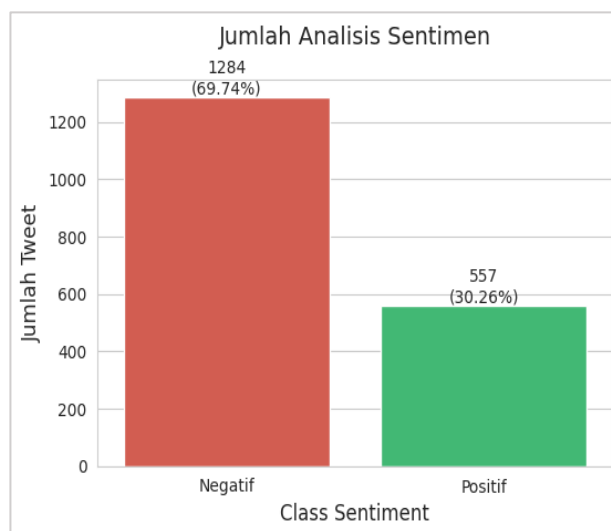
Pada saat proses pengumpulan data menghasilkan 2.116 data teks dalam bahasa Indonesia yang mencerminkan opini masyarakat terkait deepfake. Data yang diperoleh selanjutnya menjalani tahap preprocessing untuk menghasilkan data yang bersih sebelum analisis lebih lanjut dilakukan. Berikut ini distribusi jumlah dataset dapat dilihat pada tabel 1.

Tabel 1. Jumlah dataset

Kata Kunci	Data Crawling	Data Bersih
deepfake	2.116	1.841

Pelabelan Data

Tahap pelabelan data dilaksanakan untuk menambahkan kategori sentimen pada teks yang telah diproses sebelumnya. Dalam penelitian ini, data dibagi menjadi dua kategori sentimen, yaitu positif dan negatif. Dari total 1.841 data yang telah dibersihkan, terdapat 1.284 data masuk dalam kategori sentimen negatif dan 557 data termasuk dalam kategori sentimen positif. Hasil ini mengindikasikan bahwa sebagian besar orang memiliki persepsi negatif mengenai teknologi deepfake, yang mungkin disebabkan oleh meningkatnya kekhawatiran akan penyalahgunaan teknologi ini, seperti penyebaran disinformasi dan manipulasi konten digital. Berikut ini hasil distribusi sentimen dapat dilihat pada gambar 2.



Gambar 2. Distribusi sentimen

Pembobotan TF-IDF

Setelah tahap pemberian label pada data, dilakukan pembobotan nilai kata dengan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk mengubah

teks menjadi bentuk angka agar bisa diproses oleh algoritma *Machine Learning*. Metode ini mengukur seberapa sering suatu kata muncul dalam dokumen serta seberapa penting kata tersebut dalam konteks keseluruhan dokumen, sehingga kata-kata yang lebih relevan akan memiliki nilai yang lebih tinggi. Kemudian, nilai TF-IDF ini dipakai sebagai fitur dalam proses pengklasifikasian menggunakan algoritma *Support Vector Machine (SVM)*. Contoh hasil pembobotan nilai TF-IDF ditunjukkan pada tabel 2.

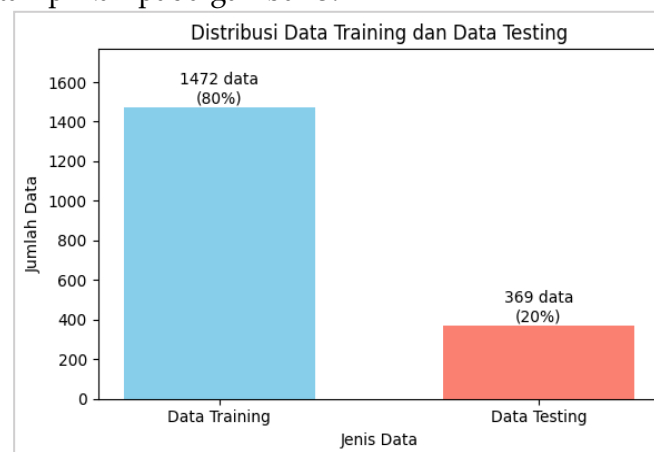
Tabel 2. Hasil pembobotan TF-IDF

Kata	Bobot
deepfake	0.060487
ai	0.035880
pakai	0.020749
video	0.019794
ya	0.018844
orang	0.017968
banget	0.016678
teknologi	0.014081
deepfake ai	0.013219
korban	0.012787

Berdasarkan Tabel 2, istilah “*deepfake*” menunjukkan bobot paling tinggi yang menandakan dominasi topik dalam dataset, diikuti oleh kata “ai”, “video”, dan “teknologi” yang mencerminkan keterkaitan pembahasan dengan aspek teknologi. Selain itu, adanya kata seperti “korban” menandakan adanya kecemasan terhadap dampak buruk dari deepfake, sehingga memperkuat penemuan bahwa pandangan masyarakat cenderung didominasi oleh persepsi negatif.

Pembagian Data (Split Data)

Setelah proses pembobotan menggunakan TF-IDF, dataset dibagi menjadi dua bagian yaitu data training dan data testing dengan proporsi 80:20. Data training digunakan untuk melatih model klasifikasi Support Vector Machine (SVM) agar dapat mengidentifikasi pola sentimen dalam data tweet, sedangkan data testing berfungsi untuk mengukur kinerja model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya. Penyebaran data training dan data testing ditampilkan pada gambar 3.



Gambar 3. Distribusi pembagian data

Penerapan Model Support Vector Machine

Pada fase ini, algoritma *Support Vector Machine* (SVM) diterapkan untuk mengkategorikan sentimen dalam data tweet. Model SVM dibangun dengan memanfaatkan library *scikit-learn* menggunakan kernel linear, yang sering dipakai dalam klasifikasi teks karena kemampuannya dalam mengelola data dengan banyak fitur. Kernel linear berfungsi dengan menemukan hyperplane yang optimal untuk memisahkan dua kategori sentimen, yakni positif dan negatif. Selain itu, parameter `class_weight = balanced` digunakan agar model dapat lebih baik dalam menangani ketidakseimbangan jumlah data di antara kategori sentimen.

Untuk meningkatkan kinerja model, dilakukan proses pengaturan hyperparameter melalui metode `GridSearchCV` dengan teknik 5-fold cross validation. Dalam proses ini, beberapa nilai parameter C diuji, yaitu 0.1, 1, dan 10. Pemilihan parameter yang terbaik didasarkan pada nilai rata-rata F1-score dari hasil cross validation, karena metrik ini memperhatikan keseimbangan antara precision dan recall. Hasil perbandingan nilai parameter C dengan F1-score akan ditampilkan pada tabel 3.

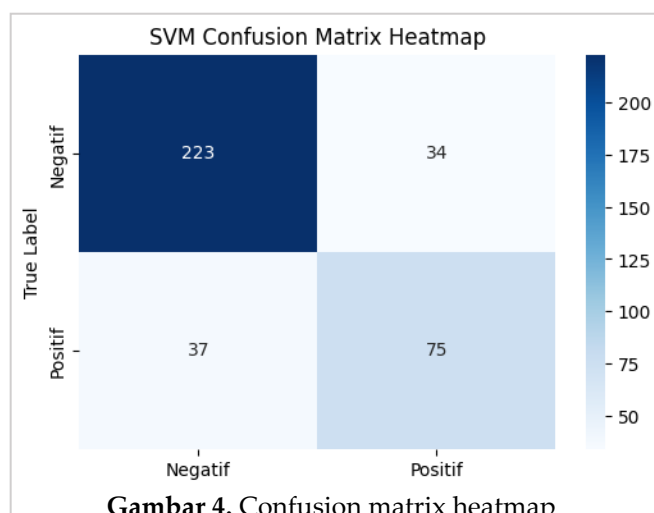
Tabel 3. Perbandingan nilai parameter C berdasarkan F1-score

C	Rata-rata F1-Score
0.1	0.518428
1	0.668854
10	0.640532

Berdasarkan hasil pengujian, nilai rata-rata F1-score tertinggi yaitu 0.668854 diperoleh dari parameter $C = 1$ jika dibandingkan dengan $C = 0.1$ dan $C = 10$. Hal ini mengindikasikan bahwa parameter tersebut dapat menyediakan keseimbangan yang ideal antara kompleksitas model dan kemampuan untuk generalisasi. Nilai C yang sangat rendah cenderung menyebabkan model underfitting, sementara nilai yang terlalu tinggi berisiko mengakibatkan overfitting. Oleh karena itu, $C = 1$ dipilih sebagai parameter terbaik untuk digunakan dalam klasifikasi model SVM.

Evaluasi Model

Confusion matrix digunakan untuk menilai kinerja model klasifikasi dengan membandingkan label sentimen yang sebenarnya dengan prediksi yang dihasilkan oleh model. Melalui confusion matrix ini dapat dilihat berapa banyak prediksi yang tepat maupun yang keliru untuk setiap kategori sentimen, sehingga memberikan gambaran yang lebih baik



Gambar 4. Confusion matrix heatmap

mengenai kemampuan model dalam melakukan klasifikasi. Dari hasil pengujian model *Support Vector Machine* (SVM) terhadap data testing, didapatkan confusion matrix yang ditampilkan pada gambar 4 berikut ini.

Berdasarkan confusion matrix yang telah didapatkan sebelumnya, tahap selanjutnya adalah mengevaluasi kinerja model dengan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan F1-score. Metrik-metrik ini dipakai untuk menilai kemampuan model dalam mengklasifikasikan sentimen secara keseluruhan serta dalam mengenali setiap kategori sentimen secara terpisah.

Tabel 4. Hasil evaluasi model svm

Metrik	Nilai
Accuracy	80.76%
Precision	68.81%
Recall	66.96%
F1-Score	67.87%

Berdasarkan hasil pada tabel 4 model *Support Vector Machine* (SVM) menunjukkan kinerja yang baik dengan tingkat akurasi 80,76%, *precision* 68,81%, *recall* 66,96%, dan F1-score 67,87%. Nilai F1-score yang berada di bawah 70% menunjukkan bahwa performa model masih berada dalam kategori sedang (moderate). Selain itu, selisih antara nilai *precision* dan *recall* menunjukkan bahwa model belum berhasil mengklasifikasikan semua data dengan baik, terutama untuk kelas sentimen positif yang memiliki tingkat kesalahan prediksi lebih tinggi dibanding kelas negatif. Selain menerapkan metrik evaluasi secara umum, kinerja model juga bisa dianalisis lebih mendalam melalui classification report yang menunjukkan nilai *precision*, *recall*, dan F1-score untuk setiap kategori sentimen, yaitu sentimen positif dan negatif. Dengan adanya laporan ini, dapat diketahui sejauh mana kemampuan model dalam mengklasifikasikan masing-masing kategori sentimen secara lebih detail. Hasil dari classification report model *Support Vector Machine* (SVM) dalam penelitian ini dapat dilihat pada tabel 5.

Tabel 5. Clasification report model SVM

Kelas Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.86	0.87	0.86	257
Positif	0.69	0.67	0.68	112
Accuracy			0.81	369
Macro Average	0.77	0.77	0.77	369
Weighted Average	0.81	0.81	0.81	369

Berdasarkan Tabel 5, model ini menunjukkan hasil yang lebih baik dalam mengidentifikasi sentimen negatif dibandingkan sentimen positif, dengan F1-score mencapai 0.86 untuk kelas negatif dan 0.68 untuk kelas positif. Ini menunjukkan bahwa ada ketidakseimbangan dalam kinerja model, dimana model cenderung lebih akurat dalam mendeteksi sentimen negatif. Perbedaan ini disebabkan oleh distribusi data yang tidak merata, di mana data sentimen negatif lebih banyak dibandingkan data sentimen positif, sehingga model lebih terlatih pada pola yang berkaitan dengan sentimen negatif. Akibatnya, kemampuan model untuk mengenali sentimen positif menjadi lebih rendah.

PEMBAHASAN

Hasil dari analisis menunjukkan bahwa pandangan masyarakat tentang teknologi *deepfake* didominasi oleh sentimen negatif, yang menandakan bahwa sebagian besar orang memiliki persepsi kritis terhadap teknologi ini. Keadaan ini mungkin disebabkan oleh meningkatnya kekhawatiran mengenai kemungkinan penyalahgunaan *deepfake*, seperti penyebaran disinformasi, manipulasi konten digital, serta risiko terhadap privasi. Hasil ini juga didukung oleh visualisasi *wordcloud* yang menampilkan dominasi kata-kata seperti “korban”, “bahaya”, dan “penipuan”, sehingga dapat disimpulkan bahwa pembicaraan mengenai *deepfake* di media sosial tidak hanya terkait dengan aspek teknologi, tetapi juga mencerminkan kekhawatiran sosial dan etika yang berkembang dalam masyarakat.

Selain itu, berdasarkan hasil evaluasi, model SVM mendapatkan akurasi sebesar 80,76% dan F1-score sebesar 67,87%, yang menandakan bahwa kinerja model berada pada kategori sedang (*moderate*). Nilai F1-score yang belum mencapai 70% menunjukkan bahwa model ini masih memiliki keterbatasan dalam mengklasifikasikan sentimen secara optimal. Laporan klasifikasi juga menunjukkan bahwa model lebih efektif dalam mengidentifikasi sentimen negatif dibandingkan dengan sentimen positif, yang menunjukkan adanya bias dalam model disebabkan oleh ketidakseimbangan distribusi data serta keterbatasan metode TF-IDF dalam memahami konteks semantik.

Jika dibandingkan dengan penelitian sebelumnya, algoritma *Support Vector Machine* (SVM) diketahui memiliki kinerja yang konsisten dalam menganalisis sentimen teks, terutama pada data berdimensi tinggi seperti teks, meskipun tidak selalu memberikan tingkat akurasi tertinggi jika dibandingkan dengan metode yang menggunakan *deep learning*. Hal ini sejalan dengan penelitian [8] yang menunjukkan bahwa metode konvensional seperti SVM masih bersaing, meskipun mulai tergeser oleh model berbasis *deep learning*. Selain itu, studi [27] menyebutkan bahwa model seperti BERT menunjukkan hasil yang lebih baik karena mampu memahami konteks bahasa dengan lebih mendalam dibandingkan dengan metode berbasis TF-IDF. Namun, SVM tetap relevan untuk digunakan karena memiliki keunggulan dalam efisiensi komputasi serta kinerja yang stabil pada dataset berukuran menengah.

Namun, penelitian ini memiliki beberapa keterbatasan, diantaranya penerapan metode berbasis leksikon dalam pelabelan data yang dapat mengakibatkan noise karena tidak memperhitungkan konteks kalimat secara keseluruhan. Selain itu, metode TF-IDF yang diterapkan hanya bergantung pada frekuensi kata tanpa melihat hubungan makna antar kata, serta ketidakseimbangan dalam distribusi data yang membuat model cenderung bias terhadap kelas mayoritas.

SIMPULAN

Berdasarkan hasil penelitian, masyarakat Indonesia menunjukkan pandangan negatif terhadap fenomena *deepfake* di platform X. Hal ini mencerminkan kekhawatiran mengenai kemungkinan penyalahgunaan teknologi, seperti penyebaran disinformasi dan manipulasi media digital. Model *Support Vector Machine* (SVM) yang diterapkan mencapai akurasi 80,76% dan F1-score 67,87%, yang menandakan bahwa hasil klasifikasinya berada pada tingkat sedang (*moderate*). Hasil ini menunjukkan bahwa pendekatan SVM dengan representasi fitur TF-IDF dapat mendeteksi pola sentimen pada teks, meskipun belum

sepenuhnya efektif dalam mengklasifikasikan semua kategori secara seimbang. Penelitian ini memiliki keterbatasan, diantaranya penggunaan pelabelan lexicon-based yang tidak memperhitungkan konteks kalimat dengan baik dan ukuran dataset yang relatif terbatas. Oleh karena itu, disarankan agar penelitian berikutnya menerapkan metode deep learning atau teknik penyeimbangan data untuk meningkatkan kemampuan model dalam analisis sentimen.

DAFTAR PUSTAKA

- [1] Q. He, S. Zhang, Y. Cai, W. Yuan, L. Ma, and K. Yu, "A Survey on Exploring Real and Virtual Social Network Rumors: State-of-the-Art and Research Challenges," *ACM Comput. Surv.*, vol. 57, no. 7, Feb. 2025, doi: 10.1145/3716498.
- [2] A. A. Khan, A. A. Laghari, S. A. Inam, S. Ullah, M. Shahzad, and D. Syed, "A survey on multimedia-enabled deepfake detection: state-of-the-art tools and techniques, emerging trends, current challenges & limitations, and future directions," *Discover Computing*, vol. 28, Apr. 2025, doi: 10.1007/s10791-025-09550-0.
- [3] G. Widjaja, "Deepfake Dan Masa Depan Kebenaran: Implikasi Etis Dan Sosial," *Berajah Journal*, vol. 5, no. 2, 2025.
- [4] "Digital 2024: Indonesia - DataReportal - Global Digital Insights." Accessed: Jan. 23, 2026. [Online]. Available: <https://datareportal.com/reports/digital-2024-indonesia>
- [5] J. A. Santiko and S. Bahri, "Analisis Wacana Pada Fenomena Penggunaan Artificial Intelligence (Ai) Dalam Konten Pemilu: Studi Kasus Konten Deepfake Soeharto Mengajak Untuk Memilih Partai Golkar Pada Media Sosial Twitter (X)," *INNOVATIVE: Journal Of Social Science Research*, vol. 4, no. 3, pp. 13215–13231, 2024.
- [6] R. Raman *et al.*, "Fake news research trends, linkages to generative artificial intelligence and sustainable development goals," *Heliyon*, vol. 10, no. 3, Jan. 2024, doi: 10.1016/j.heliyon.2024.e24727.
- [7] N. A. Sharma, A. B. M. S. Ali, and M. A. Kabir, "A review of sentiment analysis: tasks, applications, and deep learning techniques," *Int. J. Data Sci. Anal.*, Jul. 2024, doi: 10.1007/s41060-024-00594-x.
- [8] P. D. Michailidis, "A Comparative Study of Sentiment Classification Models for Greek Reviews," *Big Data and Cognitive Computing*, vol. 8, no. 9, Sep. 2024, doi: 10.3390/bdcc8090107.
- [9] J. E. Br Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 14, no. 1, pp. 42–53, Apr. 2024, doi: 10.34010/jamika.v14i1.11946.
- [10] S. A. Gulo, A. A. Pertiwi, S. P. S. Nasution, and H. Syahputra, "Deteksi Deepfake Dalam Citra Menggunakan Convolutional Neural Network (CNN)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 5, Oct. 2025.
- [11] S. A. Nugraha, "Penerapan Lexicon Based Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Danantara," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, Jun. 2025.
- [12] M. Kumar, L. Khan, and H. T. Chang, "Evolving techniques in sentiment analysis: a comprehensive review," *PeerJ Comput. Sci.*, vol. 11, Jan. 2025, doi: 10.7717/peerj-cs.2592.

- [13] S. A. Rismawan and Y. Syahidin, "Implementasi Website Berita Online Menggunakan Metode Crawling Data Dengan Bahasa Pemrograman Python," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 3, pp. 167–178, Sep. 2023.
- [14] Y. MZ, J. E. Bororing, and N. Fuadiah, "Penerapan Metode K-Nearest Neighbor Dan Decision Tree Untuk Analisis Sentimen (Studi Kasus Mario Dandi)," *Indonesian Journal of Information Technology (IJITECH)*, vol. 1, no. 1, 2023, doi: 10.71155/fcy1h865.
- [15] Stefanni, Zulfachmi, Zulkipli, and A. Saputra, "Analisis Sentimen Pengguna X Terhadap Kebocoran Data Pribadi Menggunakan Algoritma Naïve Bayes Classifier," *Jurnal Bangkit Indonesia*, vol. 14, no. 1, pp. 32–40, Mar. 2025.
- [16] M. H. Mahendra, D. T. Murdiansyah, and K. M. Lhaksana, "Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer," *Dike : Jurnal Ilmu Multidisiplin*, vol. 1, no. 2, pp. 37–43, Aug. 2023, doi: 10.69688/dike.v1i2.35.
- [17] C. V. Angkoso, M. A. N. Thrisna, B. D. Satoto, and A. Kusumaningsih, "Optimasi Klasifikasi Sentimen Menggunakan Random Forest dengan Preprocessing K-Means Clustering dan SMOTE," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 10, no. 3, Dec. 2024.
- [18] Regina, T. H. Saragih, and D. Kartini, "Analisis Sentimen Brand Ambassador BTS Terhadap Tokopedia Menggunakan Klasifikasi Bayesian Network Dengan Ekstraksi Fitur TF-IDF," *JIP (Jurnal Informatika Polinema)*, vol. 9, no. 4, Aug. 2023, doi: 10.33795/jip.v9i4.1333.
- [19] M. Al Khadafi, K. P. Kartika, and F. Febrinita, "Penerapan Metode Naïve Bayes Classifier Dan Lexicon Based Untuk Analisis Sentimen Cyberbullying Pada BPJS," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, Sep. 2022, doi: 10.36040/jati.v6i2.5633.
- [20] K. T. Putra, M. A. Hariyadi, and C. Crysdian, "Perbandingan Feature Extraction TF-IDF Dan BOW Untuk Analisis Sentimen Berbasis SVM," *Jurnal Cahaya Mandalika*, vol. 3, no. 2, pp. 1449–1463, 2022.
- [21] Y. A. Prasetyo, E. Utami, and A. Yaqin, "Pengaruh Komposisi Split Data Terhadap Performa Akurasi Analisis Sentimen Algoritma Naïve Bayes dan SVM," *Journal of Electrical Engineering and Computer (JEECOM)*, vol. 6, no. 2, 2024, doi: 10.33650/jeeecom.v4i2.
- [22] I. S. Aisah, B. Irawan, and T. Suprapti, "Algoritma Support Vector Machine (SVM) Untuk Analisis Sentimen Ulasan Aplikasi Al Qur'an Digital," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, Dec. 2023, doi: 10.36040/jati.v7i6.8263.
- [23] 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, Apr. 2023, doi: 10.34010/komputa.v12i1.9418.
- [24] F. R. Valerian, M. Syarief, and D. A. Fatah, "Klasifikasi Tingkat Obesitas Menggunakan Metode GBM Dan Confusion Matrix," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, Apr. 2025, doi: 10.36040/jati.v9i2.13062.
- [25] M. L. B. Permadi and R. Gumilang, "Penerapan Algoritma CNN (Convolutional Neural Network) Untuk Deteksi Dan Klasifikasi Target Militer Berdasarkan Citra Satelit," *Jurnal Sosial dan Teknologi (SOSTECH)*, vol. 4, no. 2, Feb. 2024, doi: 10.59188/jurnalsostech.v4i2.1138.
- [26] M. Haikal, Martanto, and U. Hayati, "Analisis Sentimen Terhadap Penggunaan Aplikasi Game Online PUBG Mobile Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, Dec. 2023, doi: 10.36040/jati.v7i6.8174.

-
- [27] J. Li, T. Tang, W. X. Zhao, J. Y. Nie, and J. R. Wen, "Pre-Trained Language Models for Text Generation: A Survey," *ACM Comput. Surv.*, vol. 56, no. 9, Apr. 2024, doi: 10.1145/3649449.