

Analisis Sentimen Pemilihan Calon Presiden Indonesia di Media Sosial Menggunakan Algoritma Support Vector

Ade Rahman^{1*}, Suhendro Yusuf Irianto²

^{1,2}Magister Teknik Informatika, Informatics & Business Institute Darmajaya

Abstract. The presidential election held every five years in Indonesia is a crucial component of the democratic process. This study focuses on the significant role of social media, particularly Twitter, as a global expression platform. Utilizing the Support Vector Machine (SVM) algorithm, the research analyzes sentiments related to the Indonesian presidential election through Twitter data. The research methodology involves data collection via Twitter crawling, data preprocessing, data partitioning, as well as the creation, training, and evaluation of the SVM model. The sentiment analysis results indicate a 97% accuracy, with 99% being negative comments. Testing on the dataset of presidential candidate Ganjar Pranowo shows SVM achieving an average accuracy of 98.61%, precision of 98.81%, and recall of 99.79%. This study underscores the effectiveness of SVM in sentiment classification on Twitter data related to presidential elections, offering valuable insights into public perspectives. The findings can be instrumental in gaining a deeper understanding of political dynamics and public opinion.

Kata kunci: Machine Learning, SVM, Data Mining, Presidential

Received Juni 2024 / Revised Juni 2024 / Accepted Juni 2024

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



PENDAHULUAN

Setiap lima tahun sekali, diadakan pemilihan presiden sebagai bagian dari proses demokrasi, terutama di negara Indonesia. Media sosial saat ini sedang ramai dibicarakan menjelang akhir pemilihan presiden [1]. Setiap pemilihan presiden tentunya akan berdampak besar bagi masyarakat dan politik nasional untuk periode selanjutnya. Oleh karena itu, menentukan pandangan masing-masing komunitas di media sosial sangat penting untuk menentukan analisis sentimen dan menentukan posisi masing-masing kandidat yang bersaing. Twitter adalah platform media sosial di mana pengguna dari seluruh dunia dapat berbagi opini dan komentar mereka. Selain menjadi sarana untuk mengekspresikan diri, Twitter juga sering digunakan sebagai sumber informasi untuk menganalisis opini publik tentang berbagai institusi dan individu [2].

Secara khusus, Twitter dapat digunakan untuk memantau perkembangan sentimen terkait pendapat seorang individu terhadap politikus yang mencalonkan diri sebagai presiden Indonesia pada 2024. Adanya kemajuan teknologi juga memungkinkan masyarakat untuk dengan mudah mengakses informasi yang diperlukan, sehingga meningkatkan kemampuan mereka untuk memahami opini politik yang ada di media sosial. Dalam hal ini analisis sentimen dapat digunakan untuk menganalisis pendapat tentang pemilihan presiden, berbagai faktor dapat diperhitungkan. Pertama, dapat dibentuk opini umum tentang visi, misi, dan program kerja calon presiden [1]. Tanggapan positif dan negatif ini dapat menunjukkan tingkat kepercayaan dan dukungan untuk calon, serta ketidaksepakatan atau keraguan terhadap calon presiden [3].

Analisis sentimen memiliki manfaat yang penting dalam pengembangan sistem untuk menganalisis, mengidentifikasi, dan mengungkapkan pendapat. Manfaat dari text analytics adalah memungkinkan sistem untuk melakukan analisis, identifikasi, dan ekspresi pendapat dalam bentuk teks. Proses ini melibatkan identifikasi sentimen atau opini dari teks yang terdapat dalam dataset. Teks yang berhubungan dengan topik atau peristiwa positif maupun negatif tersedia di internet melalui berbagai media seperti blog, forum, media sosial, serta situs ulasan. Meskipun sentimen opini dapat diproses secara manual, semakin banyak sumber

¹*Korespondensi author.

Alamat Email: ade.2221210001@mail.darmajaya.ac.id (rahman), suhendro@darmajaya.ac.id(irianto)

opini yang diperoleh akan semakin membutuhkan waktu dan tenaga dalam mengklasifikasikan polaritas opini tersebut.

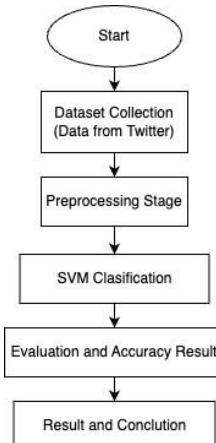
Algoritma klasifikasi juga memiliki kelebihan dan kekurangan dalam mengklasifikasikan teks. Algoritma Support Vector Machine memiliki akurasi dan generalitas klasifikasi yang tinggi [4]. SVM (Support Vector Machine) adalah metode untuk analisis data [4], [5]. Algoritma bekerja untuk klasifikasi dan juga untuk analisis regresi. Hal ini bertujuan untuk meningkatkan ukuran kinerja dalam menggunakan Algoritma SVM (Support Vector Machine) serta mampu menganalisis hasil opini terhadap analisis sentimen media sosial calon presiden 2024 [6].

Pada penelitian sebelumnya, analisis sentimen Pilpres 2019 telah memberikan kontribusi yang berharga pada analisis sentimen dengan menggunakan metode SVM. pemahaman yang lebih baik tentang penggunaan metode ini dalam mengklasifikasikan sentimen teks dapat diperoleh. Dalam analisis sentimen mengenai calon Presiden RI 2019 dari media sosial Twitter, metode analisis sentimen dengan algoritma klasifikasi SVM capres Indonesia 2019 dengan fungsi Kernel dapat mengklasifikasikan dengan akurasi sebesar 86,82% untuk tweet dengan kata kunci “Jokowi” dan 86,27% untuk tweet dengan kata kunci “Prabowo” [7].

Hasil yang diperoleh dapat dijadikan pijakan untuk melanjutkan penelitian lebih lanjut dan menganalisis sentimen dengan akurasi yang lebih tinggi. Dengan mempelajari temuan penelitian sebelumnya, dapat diketahui kekuatan dan keterbatasan kedua metode tersebut, serta mengoptimalkan parameter yang digunakan untuk mencapai hasil yang lebih baik.

METODE

Metode penelitian ini melibatkan beberapa tahapan, dimulai dengan pengumpulan data seperti yang digambarkan pada *Gambar 1*. Pengumpulan data dilakukan melalui proses crawling data Twitter. Selanjutnya tahap pengolahan data dilakukan pada proses preprocessing, kemudian pembagian data, pembuatan model, pelatihan model, evaluasi model, penggunaan model [8].



Gambar Tahapan Penelitian

Berikut ini merupakan penjelasan dari gambar yang merupakan alur dari penelitian ini:

1. Pengumpulan data
Pada tahap ini dilakukan pengumpulan data dari twitter terkait pemilihan calon presiden Indonesia. Data tersebut mencakup sentimen terkait calon presiden, seperti positif, negatif, atau netral [9].
2. Preprocessing
Tahap awal dari Preprocessing adalah Cleansing, yang mana merupakan sebuah proses pembersihan data teks dengan menghilangkan data yang tidak relevan dan tidak konsisten [10].
3. Pembuatan Model
Setelah *preprocessing* data, langkah selanjutnya adalah membangun model Support Vector Machine (SVM). Pada model Support Vector Machine (SVM), konstruksi vektor dilakukan dengan unigram, bigram, dan trigram menggunakan TF-IDF untuk fitur data. Kemudian melakukan prediksi pada data training dan data testing, selanjutnya menetapkan nilai Akurasi ke nilai yang lebih tinggi untuk menghasilkan overfitting, mengevaluasi model pada data training dan data testing, kemudian menampilkan hasil evaluasi pada data training dan data testing serta memberikan saran.

SVM (Support Vector Machine) adalah sebuah model yang berasal dari teori belajar statistik dan telah terbukti memberikan hasil yang lebih baik dibandingkan dengan metode lainnya seperti yang disebutkan oleh Sulistiani pada tahun 2018. Karena kemampuannya tersebut, SVM sering digunakan dalam pengelompokan label yang telah didapatkan dari proses pembobotan dan pelabelan sehingga menghasilkan analisis sentimen. Dalam konteks analisis sentimen, SVM bisa digunakan untuk mengklasifikasikan teks menjadi tiga kelas, yaitu positif, negatif, dan netral.

Model SVM akan mempelajari pola dan hubungan antar teks fitur dan label sentimen yang terkait. Oleh menggunakan metode pembelajaran statistik yang kuat, SVM dapat memisahkan dan mengklasifikasikan teks menjadi tiga yang berbeda kelas sentimen. Pada tahap klasifikasi, SVM akan memanfaatkan hasil pembobotan dan label sentimen yang dimiliki telah diberikan pada data. SVM akan menggunakan ini informasi untuk melatih model dan menemukan yang optimal hyperplane yang dapat memisahkan teks dengan positif, negatif, dan sentimen netral. Hasilnya adalah sebuah Model SVM yang dapat digunakan untuk mengklasifikasikan teks baru menjadi salah satu dari tiga kelas sentimen [11]. Dengan menggunakan SVM dalam mengelompokkan sentimen ke dalam tiga kelas tersebut, kita bisa mendapatkan pemahaman yang lebih dalam tentang sentimen terkandung dalam teks. Informasi tentang apakah teks memiliki sentimen positif, negatif, atau netral berguna dalam berbagai aplikasi, seperti pelanggan analisis opini, pemantauan media sosial, atau pengguna pendataan umpan balik.

Dengan demikian, SVM memberikan kontribusi yang signifikan dalam mengelompokkan label sentimen dalam tiga kelas, yaitu positif, negatif, dan netral, dan membantu lebih baik memahami sentimen dalam teks.

4. Evaluasi

Pada tahap ini dilakukan pengujian terhadap algoritma yang digunakan untuk mendapatkan informasi model yang akurat. Pengukuran klasifikasi algoritma *Support Vector Machine*. Untuk Analisis Sentimen Pemilihan Calon Presiden Indonesia di Media Sosial Menggunakan Algoritma Support Vector Machine

HASIL DAN DISKUSI

Dalam penelitian ini peneliti melakukan empat tahapan dalam pelaksanaan penelitian yaitu Pengumpulan Data, *Preprocessing*, Pembuatan Model, Evaluasi.

1. Pengumpulan Data

Data tweet twitter diperoleh dari hasil komentar sebanyak 3.022 tweet dengan kata kunci Ganjar Pranowo, Anies Baswedan dan Prabowo Subianto. Data ini berupa tweet yang berisi kata atau frase terkait calon presiden. memanfaatkan fasilitas twitterscraper dengan menginstal node js untuk desktop. Cara ini dapat memudahkan proses crawling, karena dalam proses crawling tidak ada batasan tweet dan hasilnya bisa langsung disimpan dalam bentuk file .json atau .csv kemudian disimpan dalam file excel.

Tabel 1. Contoh hasil crawling data twitter

Username	Tweet
leogusta_	@BeBuzzerNKRI @dwioktariyadi @aniesbaswedan Selama bukan mengolok-ngolok, fine-fine aja menurut gw. Prabowo juga bikin pin Garuda Merah dulu. Selain itu, kaos-kaos yg dipake masyarakat sipil dengan logo burung garuda, banyak kreativitasnya. Acara pemerintahanpun sama.
richworld	@jansen_jsp @aniesbaswedan @AgusYudhoyono Ditunggu kejutannya ya pak. Semoga berhasil mengagetkan kita semua. Jangan2 calon wakilnya anies adalah pak @jansen_jsp. \nðŸ˜,

BEEQUEN2w	@DaengWahidin2 @DokterTifa @aniesbaswedan Hihi aku ga bakal pilih kalau mau nyalon Gub .DKI ..belum nyampai ...eee belanda masih jauh.
ProoFaris29119	@spesies_galaxy @PartaiSocmed @aniesbaswedan @susipudjiastuti ðŸ˜ ðŸ˜ benr
BiruScooter	@MANWAR2711 @rh_dewa @RaknongPem @msobri99 @AlineRaztha @aniesbaswedan Itu dah dikasih contoh msh tanya lg..ckckck

2. Preprocessing

Fungsi dari preprocessing adalah untuk menghapus karakter-karakter yang tidak penting, seperti hashtag (#), angka, username (@), URL, tanda baca, dan emotikon. Selain itu, proses Case folding dilakukan untuk mengubah bentuk kata dari karakter-karakter tersebut agar seragam (lower case). Selanjutnya, Tahap Tokenize adalah proses untuk memecah teks menjadi unit-unit yang lebih kecil, seperti kata atau frase. dari tahap tokenize diperoleh representasi teks yang lebih terstruktur untuk dianalisis [12]. Dan tahapan Remove Stopwords pada analisis sentimen ini merupakan proses menghilangkan stopwords atau kata-kata umum yang tidak memberikan banyak informasi pada analisis sentimen. Stopword umumnya terdiri dari kata-kata seperti "dan", "atau", "dari", "di", "ke", dan seterusnya. Tahapan *preprocessing* ini mengacu pada penelitian yang dilakukan oleh allif rizki Abdillah dan Firman Noor Hasan 2022 pada [13].

3. Labeling

Proses selanjutnya adalah pemberian label. Proses ini dilakukan untuk melabeli isi komentar pengguna twitter yang selanjutnya akan disimpan dalam dataset untuk proses Pemisahan Data. Ada tiga pilihan dalam pelabelan kalimat yaitu Positif, Netral dan Negatif, hal ini dimaksudkan agar pemilihan label konsisten dan menghindari input yang salah.

Tabel 2. Pelabelan data

Username	Tweet	Sentiment
leogusta_	@BeBuzzerNKRI @dwioktariyadi @aniesbaswedan Selama bukan mengolok-ngolok, fine-fine aja menurut gw. Prabowo juga bikin pin Garuda Merah dulu. Selain itu, kaos-kaos yg dipake masyarakat sipil dengan logo burung garuda, banyak kreativitasnya. Acara pemerintahanpun sama.	negative
richworld	@jansen_jsp @aniesbaswedan @AgusYudhoyono Ditunggu kejutannya ya pak. Semoga berhasil mengagetkan kita semua. Jangan2 calon wakilnya anies adalah pak @jansen_jsp. \nðŸ˜,	negative
BEEQUEN2w	@DaengWahidin2 @DokterTifa @aniesbaswedan Hihi aku ga bakal pilih kalau mau nyalon Gub .DKI ..belum nyampai ...eee belanda masih jauh.	negative
ProoFaris29119	@spesies_galaxy @PartaiSocmed @aniesbaswedan @susipudjiastuti ðŸ˜ ðŸ˜ benr	negative
BiruScooter	@MANWAR2711 @rh_dewa @RaknongPem @msobri99 @AlineRaztha @aniesbaswedan Itu dah dikasih contoh msh tanya lg..ckckck	negative

4. Model Klasifikasi

Setelah data melalui proses preprocessing dan pelabelan, data tersebut kemudian dianalisis menggunakan algoritma SVM. Proses ini menggunakan teknik data mining dan model klasifikasi yang digunakan dalam penelitian ini adalah Support Vector Machine (SVM). Pengujian dilakukan dengan bantuan ekstensi Google Colab untuk mengetahui Accuracy, Precision dan recall. Dengan pengujian tersebut maka akan mengetahui opini publik apakah cenderung ke positif, negative atau netral [14].

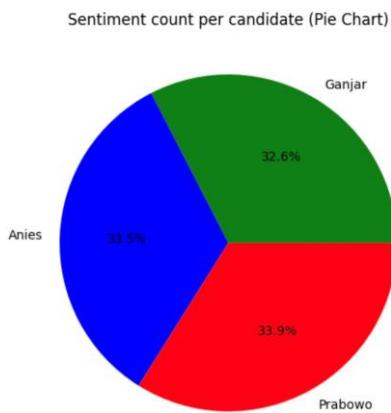
5. Evaluasi

Berdasarkan hasil data yang dihasilkan dari klasifikasi Support Vector Machine, didapatkan nilai akurasi sebesar 97%. Yang terdiri dari 99% negatif dan 1% positif. Hal ini menunjukkan bahwa analisis sentimen dengan metode SVM menghasilkan crawling data dari tweet yang banyak menampilkan komentar negatif terkait topik pemilihan calon presiden Indonesia. Untuk mengevaluasi akurasi suatu model, diperlukan suatu pengukuran evaluasi yang disebut confusion matrix. Confusion matrix merupakan tabel yang menggambarkan sejumlah besar data pengujian yang diprediksi secara kuat maupun lemah oleh model klasifikasi yang digunakan. Confusion matrix dapat memberikan informasi penting terkait performa model klasifikasi, seperti recall, precision, dan accuracy [15]. Dalam konteks SVM, berikut ini adalah hasil confusion matrix yang dihasilkan.

Akurasi: 0.9752066115702479				
Laporan Klasifikasi:				
	precision	recall	f1-score	support
negative	0.98	1.00	0.99	589
positive	1.00	0.06	0.12	16
accuracy			0.98	605
macro avg	0.99	0.53	0.55	605
weighted avg	0.98	0.98	0.96	605
Classification Report:				
	precision	recall	f1-score	support
negative	0.98	1.00	0.99	589
positive	1.00	0.06	0.12	16
accuracy			0.98	605
macro avg	0.99	0.53	0.55	605
weighted avg	0.98	0.98	0.96	605

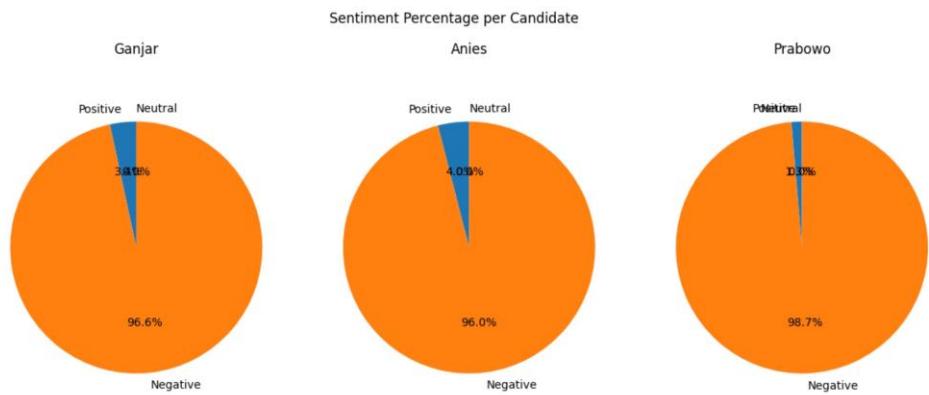
Gambar 2. Laporan hasil akurasi model svm

Hasil dari klasifikasi SVM telah disajikan dalam format grafik di bawah ini. Berdasarkan pengujian terhadap dataset Ganjar Pranowo, algoritma SVM berhasil mencapai rata-rata akurasi sebesar 98,61%, presisi sebesar 98,81%, dan recall sebesar 99,79%. Grafik tersebut menunjukkan proporsi sentimen positif dan negatif terhadap kedua kandidat presiden.



Gambar 3. Proporsi sentimen terhadap kedua kandidat presiden

Terlihat di sini bahwa jumlah sentimen terbanyak adalah tweet tentang Prabowo, disusul oleh Anies dan terakhir ditempati oleh Ganjar.



Gambar 4. Persentase sentimen masing-masing kandidat

Berdasarkan analisis sentimen terhadap data yang ada, diperoleh hasil sebagai berikut:

1. Kandidat Ganjar memiliki sentimen negatif sebesar 96,6% dengan sentimen positif sebesar 3,4%.
2. Calon Anies memiliki sentimen negatif sebesar 96,0% dengan sentimen positif sebesar 4,0%.
3. Calon Prabowo memiliki sentimen negatif sebesar 98,7% dengan sentimen positif sebesar 1,3%.

Dari hasil tersebut dapat disimpulkan bahwa mayoritas sentimen terhadap calon Ganjar, Anies, dan Prabowo adalah negatif. Sentimen positif hanya muncul dalam proporsi yang sangat kecil. Hal ini menunjukkan bahwa ketidakpuasan atau kritik lebih dominan dalam hal penilaian terhadap ketiga capres tersebut.

KESIMPULAN

Kesimpulan penelitian dengan judul analisis sentimen calon presiden Indonesia 2024 dari media sosial twitter dengan menggunakan metode Support Vector Machine dapat memprediksi kelas sentimen capres Indonesia 2024 dari media sosial twitter dengan pembagian data latih dan data uji terbaik adalah 55% data latih dengan sebaran data latih pada masing-masing kelas negatif, positif, netral sebanyak 1600 data kemudian data uji berisi 1400 data. Pada penelitian ini Analisis sentimen calon presiden Indonesia 2024 dari media sosial twitter dengan menggunakan metode Support Vector Machine mendapatkan nilai akurasi sebesar 97%, menunjukkan bahwa metode Support Vector Machine merupakan algoritma yang baik dalam model klasifikasi.

REFERENSI

- [1] U. Rauta, "Mengagas Pemilihan Presiden yang Demokratis dan Aspiratif".
- [2] A. A. Amrullah *et al.*, *Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers Unisbank (Sendi_U) Ke-2 Tahun 2016 Kajian Multi Disiplin Ilmu dalam Pengembangan Ipteks untuk Mewujudkan Pembangunan Nasional Semesta Berencana (Pnsb) sebagai Upaya Meningkatkan Daya Saing Global Review Atas Analisis Sentimen Pada Twitter Sebagai Representasi Opini Publik Terhadap Bakal Calon Pemimpin.*
- [3] A. F. Sabil, P. P. Adikara, and M. A. Fauzi, "Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy".
- [4] Luthfanida, "Analisis Sentimen Data Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine (Svm) Tentang Presiden Jokowi 3 Periode)."'

- [5] A. N. Salsabila and Y. Afrillia, “Seminar Nasional Fakultas Teknik Universitas Malikussaleh Tahun 2022.”
- [6] T. Sinta Peringkat, berdasarkan S. Dirjen Penguanan RisBang Kemenristekdikti, A. Mustopa, A. Yadi Kuntoro, and P. Studi Ilmu Komputer, “Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa”, [Online]. Available: www.bsi.ac.id
- [7] S. Nurul, J. Fitriyyah, N. Safriadi, E. Esyudha, and P. #3, “JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes”, [Online]. Available: <http://dev.twitter.com>.
- [8] C. Intelligence and Neuroscience, “Retracted: Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques,” *Comput Intell Neurosci*, vol. 2023, pp. 1–1, Oct. 2023, doi: 10.1155/2023/9810910.
- [9] M. H. Al-Areef and K. Saputra, “Nomor 2,” *Agustus*, vol. 22, pp. 270–279, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jis/index>
- [10] Asno Azzawagma Firdaus, Anton Yudhana, and Imam Riadi, “Analisis Sentimen Pada Proyeksi Pemilihan Presiden 2024 Menggunakan Metode Support Vector Machine,” *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 236–245, Jun. 2023, doi: 10.51454/decode.v3i2.172.
- [11] Hayatin Nur, G. I. Marthasari, and Nuraini Lia, “Optimization of Sentiment Anaysis for Indonesian Presidential Election using Naive Bayes and Particle Swarm Optimization”.
- [12] V. R. Hananto, “Analisis Sentimen Untuk Pemeringkatan Popularitas Situs Belanja Online Di Indonesia Menggunakan Metode Naive Bayes (Studi Kasus Data Sekunder).” [Online]. Available: <https://www.researchgate.net/publication/360077700>
- [13] A. R. Abdillah and F. N. Hasan, “Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier,” *Smatika Jurnal*, vol. 13, no. 01, pp. 117–130, Jul. 2023, doi: 10.32664/smatika.v13i01.750.
- [14] N. Alexander, R. Bria, and A. Witanti, “Analisis Sentimen Masyarakat Indonesia Menggunakan Algoritma Support Vector Machine Tentang Pilpres 2024,” 2023.
- [15] S. Keputusan Direktur Jenderal Pendidikan Tinggi *et al.*, “Terakreditasi Sinta Peringkat 3 Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma Svm Pada Media Sosial Twitter,” 2026.