

Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM

Arfina Handayani*, Ilka Zufria

Fakultas Sains dan Komputer, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan
Jl. William Iskandar Ps. V, Medan Estate, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara, Indonesia

Email : ^{1,*}arfina.zhaa11@email.com, ²ilkazufria@uinsu.ac.id

Email Penulis Korespondensi: arfina.zhaa11@email.com

Submitted: 07/10/2023; Accepted: 18/10/2023; Published: 23/10/2023

Abstrak—Pada tahun 2024, Indonesia akan mengadakan pemilihan kepala negara yang menjadi puncak demokrasi dalam sistem politiknya. Dalam konteks demokrasi, partisipasi aktif masyarakat memegang peran kunci dalam menjadikan pemilihan umum sukses. Meskipun demikian, masih ada kecenderungan masyarakat untuk hanya mengikuti prosedur pemilihan tanpa berpartisipasi aktif. Oleh karena itu, penelitian ini dilakukan untuk menganalisis sentimen masyarakat terhadap bakal calon presiden 2024, menggunakan data dari media sosial Twitter. Media sosial, khususnya Twitter, telah menjadi platform utama bagi masyarakat Indonesia dalam menyampaikan opini, pandangan, dan preferensi politik mereka. Dengan jumlah pengguna media sosial yang mencapai 170 juta pada tahun 2021, Indonesia memiliki potensi besar untuk mengumpulkan beragam pendapat dan sentimen terkait pemilihan umum. Penelitian ini bertujuan untuk menganalisis sentimen dengan menggunakan metode Support Vector Machine (SVM) terhadap tweet yang mengandung hashtag terkait tiga bakal calon presiden, yaitu Aniesbaswedan, Ganjarpranowo, dan PrabowoSubianto. Hasil analisis sentimen menunjukkan bahwa metode SVM dapat digunakan untuk menganalisis sentimen dengan akurasi sekitar 78.3% secara keseluruhan. Akurasi ini dipengaruhi oleh faktor-faktor seperti jumlah data yang digunakan dan komposisi data positif dan negatif. Data sentimen yang dianalisis berasal dari 1719 tweet pengguna Twitter, dengan distribusi sekitar 597 data untuk Aniesbaswedan, 627 data untuk Ganjarpranowo, dan 495 data untuk Prabowo Subianto. Penelitian ini memberikan wawasan tentang bagaimana masyarakat Indonesia menyampaikan sentimen mereka terkait bakal calon presiden 2024 melalui media sosial. Analisis sentimen ini dapat menjadi referensi penting dalam memahami preferensi politik publik menjelang pemilihan umum yang akan datang.

Kata Kunci: Sentimen; Presiden; Twitter; SVM; Indonesia

Abstract—In 2024, Indonesia will hold an election for head of state which will be the pinnacle of democracy in its political system. In the context of democracy, active community participation plays a key role in making general elections successful. However, there is still a tendency for people to only follow election procedures without actively participating. Therefore, this research was conducted to analyze public sentiment towards the 2024 presidential candidates, using data from social media Twitter. Social media, especially Twitter, has become the main platform for Indonesian people to express their political opinions, views and preferences. With the number of social media users reaching 170 million in 2021, Indonesia has great potential to gather various opinions and sentiments regarding general elections. This research aims to analyze sentiment using the Support Vector Machine (SVM) method on tweets containing hashtags related to three presidential candidates, namely Aniesbaswedan, Ganjarpranowo, and PrabowoSubianto. The results of sentiment analysis show that the SVM method can be used to analyze sentiment with an accuracy of around 78.3 % overall. This accuracy is influenced by factors such as the amount of data used and the composition of positive and negative data. The sentiment data analyzed came from 1719 tweets from Twitter users, with a distribution of around 597 data for Aniesbaswedan, 627 data for Ganjarpranowo, and 495 data for Prabowo Subianto. This research provides insight into how Indonesian people convey their sentiments regarding the 2024 presidential candidates via social media. . This sentiment analysis can be an important reference in understanding the public's political preferences ahead of the upcoming general election.

Keywords: Sentiment; President; Twitter; SVM; Indonesia

1. PENDAHULUAN

Pesta demokrasi akan diadakan di Indonesia pada tahun 2024 untuk memilih pemimpin negara. Pemilihan umum seringkali merupakan hasil dari konsep atau gagasan besar demokrasi, keadilan, dan kesejahteraan bagi setiap orang dalam suatu bangsa di segala bidang kehidupan[1]. Keberhasilan penyelenggaraan pemilu tergantung pada partisipasi masyarakat[2]. Masyarakat luas diberikan akses dan ruang untuk berpartisipasi aktif dalam pemilihan umum. Meskipun demikian, terkadang individu tetap mengikuti aturan dan tren yang adil[3]. Oleh karena itu, pemilihan umum diselenggarakan setiap lima tahun sekali secara langsung, bebas, jujur, dan rahasia[4]. Sistem demokrasi ditegakkan di Indonesia. Suara terbanyak diperlukan untuk pemilihan umum presiden, wakil rakyat, gubernur, bupati, dan kepala desa[5].

Setiap calon kepala negara akan mempertimbangkan popularitasnya, tergantung pada pendapat umum[6]. Dahulu orang mengkomunikasikan ide-ide mereka melalui media cetak, namun tidak semua orang memiliki keterampilan menulis atau akses terhadap peluang penerbitan[7]. Namun seiring kemajuan teknologi komunikasi, cara masyarakat mengkomunikasikan pandangan dan keyakinan mereka pun berubah. Media sosial telah berkembang menjadi platform bagi individu di seluruh dunia untuk berbagi informasi[8]. Orang-orang ini berbicara tentang peristiwa terkini dan mempublikasikan beragam perspektif tentang berbagai subjek. Sejak partai politik membeberkan nama-nama bakal calon Pilpres 2024, akan ramai diperbincangkan, khususnya di media sosial[9]. Oleh karena itu, penelitian ini dikembangkan sebagai landasan bagi peneliti untuk dapat menilai tingkat



sikap terhadap calon presiden 2024 berdasarkan banyaknya komentar masyarakat terhadap calon tersebut[10]. Twitter adalah platform media sosial populer yang digunakan orang untuk mengekspresikan ide-ide mereka[11]. Pengguna media sosial, khususnya pengguna Twitter di Indonesia, belakangan ini berperan penting dalam membentuk sikap, pandangan, perasaan, dan preferensi politik masyarakat, khususnya menjelang pemilu[12].

Jumlah pengguna media sosial di Indonesia saat ini mencapai 170 juta pada awal tahun 2021, dengan Twitter menjadi platform terpopuler[13]. Sebuah survei yang mengungkap posisi Indonesia dalam peringkat penggunaan Twitter telah dilakukan dan mengungkapkan bahwa masyarakat Indonesia berada di peringkat keenam secara keseluruhan. Ambang batas sebesar 63,6% dicapai untuk rentang usia 16–64 tahun, sehingga kelompok tersebut berhak memilih dalam pemilihan umum[14]. Beragamnya pemikiran yang disuarakan oleh netizen Indonesia, khususnya di Twitter, akan menimbulkan beragam tanggapan. Analisis sentimen harus dilakukan untuk menemukan dan mengidentifikasi pola postingan pengguna Twitter[15]. Proses menganalisis orang-orang yang membagikan pemikirannya tentang berbagai isu di media sosial dikenal sebagai analisis sentimen. Analisis sentimen telah digunakan secara luas, misalnya untuk menentukan bagaimana perasaan konsumen terhadap suatu produk dan kecenderungan politik mereka[16].

Berdasarkan hal tersebut di atas, analisis sentimen ini dikembangkan untuk mengidentifikasi tweet yang diposting di timeline Twitter yang berisi komentar positif dan negatif terhadap sejumlah individu yang dicalonkan oleh partai politik masing-masing, seperti partai politik PDI Perjuangan, yang telah mengumumkan akan mencalonkan Ganjar Pranowo sebagai calon. Capres 2024, Parpol Gerindra menyatakan Prabowo Subianto bakal maju sebagai capres 2024, dan Parpol Nasdem mengusung Anies Baswedan sebagai capres 2024.

Support Vector Machine (SVM) adalah teknik klasifikasi yang digunakan dalam penelitian ini untuk mengkategorikan calon presiden 2024. Metode SVM merupakan salah satu teknik yang sering digunakan untuk mengklasifikasikan data, khususnya data teks. Karena prosedur penghitungan vektor dukungan dapat dinyatakan sebagai masalah QP, salah satu keuntungannya adalah prosedur ini dapat dijalankan dengan cukup cepat[17]. Analisis sentimen dan penambangan opini merupakan disiplin akademis yang mengkaji bagaimana sentimen, opini, penilaian, sikap, dan perasaan seseorang diungkapkan dalam bahasa tertulis[18].

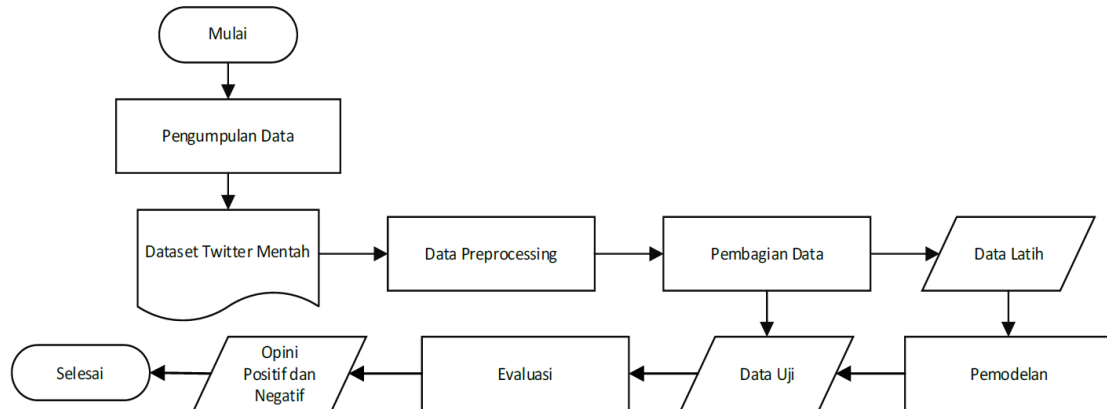
Terdapat penelitian sebelumnya yang membahas tentang Analisis Sentimen Pemilihan Presiden Indonesia 2019 di Twitter Berdasarkan Geolocation Menggunakan Metode Naïve Bayesian Classification yang menghasilkan perolehan sentimen tweet berdasarkan wilayah didapat melalui titik koordinat geografis bumi yang terhubung pada aplikasi twitter melalui API Geolocation yang terdiri dari garis lintang dan garis bujur pada google maps yang ditetapkan oleh pengguna. Perbandingan klasifikasi algoritma naïve bayesian classisfier dengan algoritma Tree.J48 memiliki tingkat akurasi yang cukup berbeda. Dari 839 data dimana klasifikasi algoritma naïve bayes classifier lebih tinggi dengan tingkat akurasi 77,62% dibandingkan dengan algoritma Tree.J48 yang hanya 70%[19]. Kemudian terdapat juga penelitian terdahulu tentang Analisis Sentimen Terhadap Isu Politik Calon Presiden Indonesia 2019 Berdasarkan Opini Netizen Dari Twitter Menggunakan Metode Klasifikasi svm yang mendapatkan calon presiden Jokowi mengalami sentiment negatif lebih besar dari pada sentiment positif. Tweet negatif sebanyak 387 dari 742 tweet terhadap Jokowi, sedangkan team Prabowo mendapatkan sentiment positif lebih banyak. Diketahui Tweet positif kepada Prabowo sebesar 273 tweet dari total 475. Dan hasil klasifikasi menggunakan svm mendapatkan akurasi svm sebesar 0.98% atau 98% terhadap Jokowi, 0,99% atau 99% akurasi svm terhadap Prabowo. Ada juga penelitian mengenai Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analisis yang mendapatkan sentimen yang di tunjukkan publik dunia maya pada saat debat berlangsung cenderung bersentimen positif dari 15.000 percakapan pada saat debat berlangsung. Berdasarkan hasil pada tabel dan barplot pada program R, hasil presentase, baik tweet yang menggunakan hastag #JokowiAminMenangDebatmaupunhastag#PrabowoIndonesiaMenang menunjukkan cenderung mengandung sentimen positif lebih banyak dari pada negatif dan netral pada setiap tweet. Untuk sentimen negatif pada kedua hastag, jumlahnya cukup banyak pada setiap tweet yang diposting saat debat berlangsung, walaupun memang tidak sebanyak sentimen positif. Pada 2019, terdapat juga penelitian mengenai Analisis Sentimen Berbasis Leksikon InSet Terhadap Paratai Politik Pserta Pemilu 2019 Pada Media Sosial Dari 839 data dimana klasifikasi algoritma naïve bayesian classifier lebih tinggi dengan tingkat akurasi 77,62% dibandingkan dengan algoritma Tree.J48 yang hanya 70%. Dan sebagai referensi terakhir terdapat penelitian mengenai Analisis Sentimen Calon Presiden Indonesia 2019 Dari Media Sosial Twitter Menggunakan Metode Naïve Bayes yang menggunakan 1500 data mendapatkan nilai akurasi yaitu 64,6% untuk paslon 01, dan 58% untuk paslon 02 pada pengujian 3 kelas. Sedangkan pengujian 2 kelas 77,7% untuk paslon 01 dan 84 pada paslon 02[20]. Berdasarkan penelitian diatas tersebut maka penelitian ini dibuat dengan perbedaan yaitu memiliki data sebanyak 1719 data, kemuadian akan dianalisis bagaimana tingkat akurasi menggunakan metode SVM dengan harapan agar memperoleh hasil akurasi yang lebih tinggi dibandingkan penelitian sebelumnya.

2. METODOLOGI PENELITIAN

2.1 Diagram Alur Penelitian

Fakta yang dikumpulkan melalui penelitian disebut temuan penelitian. Dalam hal ini yang menjadi subjek kajian adalah informasi dari tweet calon presiden Indonesia 2024, yakni yang berhashtag #aniesbaswedan,

#ganjarpranowo, dan #prabowo. Perencanaan diperlukan dalam pelaksanaan penelitian ini agar prosedur dapat berjalan sesuai dengan yang diharapkan. Proses penelitian ditampilkan dalam bentuk diagram alir pada gambar di bawah ini.



Gambar 1. Diagram Alur Penelitian

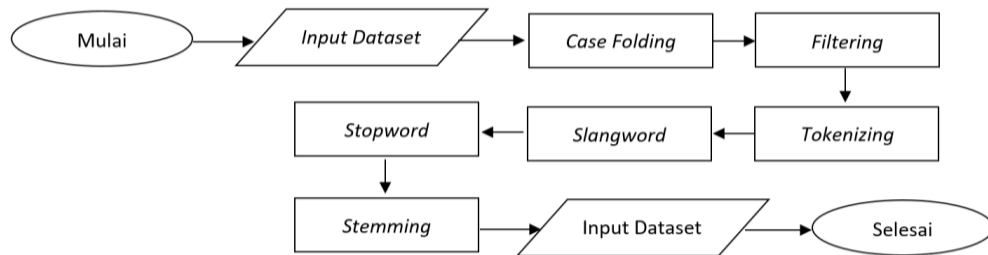
Gambar 1. Tahapan pengumpulan data, preprocessing, vektorisasi, splitting data, pemodelan klasifikasi, dan evaluasi merupakan tahapan penelitian yang akan dilakukan.

2.2 Pengumpulan Data

library Tweepy bahasa pemrograman Python digunakan untuk crawling API Twitter selama langkah pengumpulan data. Tweet pengguna dari 13 Agustus 2022 hingga 13 Agustus 2023 yang berisi tagar #aniesbaswedan, #ganjarpranowo, dan #prabowosubianto crawl untuk diambil datanya. Untuk diproses lebih lanjut, dataset yang crawling disimpan dalam format .csv.

2.3 Preprocessing Data

Dataset tersebut ditambahkan ke kategori data tidak terstruktur mengikuti prosedur pengumpulan data. Kumpulan data menjalani prapemrosesan teks sebelum analisis lebih lanjut untuk menghilangkan dan mengatasi data yang berisik serta memberikan hasil komputasi terbaik. case folding, filtering, tokenizing, dan stopwords removal semuanya disertakan dalam alur proses persiapan teks. Pelabelan data dilakukan secara manual setelah dibersihkan. Diagram alir Gambar 2. untuk persiapan teks disajikan di bawah ini.



Gambar 2. Preprocessing Data

2.4 Vektorisasi TF-IDF

Untuk memberikan kata (term) berdasarkan nilai suatu dokumen atau tweet, penelitian ini menggunakan teknik pembobotan kata yang disebut TF-IDF (Term Frequency-Inverse Document Frequency) yang berguna untuk menimbang. Persamaan (1) memberikan rumus algoritma TF-IDF sebagai berikut:

$$w_{ij} = tf_{ij} \times \log \left(\frac{D+1}{df_i+1} \right) + 1 \tag{1}$$

Keterangan :

w_{ij} = bobot kata t terhadap dokumen

tf_{ij} = jumlah kemunculan kata t_i

D = jumlah dokumen

df_{ij} = jumlah kemunculan kata pada D_j

2.5 Support Vector Machine (SVM)

Pada tahap ini, data dibagi menjadi dua kelompok: data pelatihan (80%) dan data pengujian (20%). Data pengujian digunakan sebagai tolok ukur dan untuk mengevaluasi kinerja algoritma atau model yang telah diajarkan



sebelumnya, sedangkan data pelatihan digunakan untuk melatih algoritma dan membangun pemodelan. Prosedur pemodelan SVM adalah langkah selanjutnya.

Dengan memanfaatkan fungsi kernel untuk melakukan modifikasi matematis pada ruang pembelajaran, pendekatan SVM memungkinkan komputasi untuk masalah linier. Ide utama dalam pendekatan SVM untuk mengkategorikan data adalah memilih hyperplane optimal untuk memberikan pemisahan atau jarak antara dua kelas yang telah ditetapkan. Menemukan hyperplane ideal yang memberikan pemisahan atau jarak antara dua kelas adalah cara SVM beroperasi. hyperplane dengan margin terbesar.

Tahapan dalam metode Support Vector Machine adalah sebagai berikut:

- a. Menentukan kata yang sering muncul dari tiap dokumen atau tweet yang digunakan.
- b. Menentukan inisialisasi awal untuk nilai $\alpha=0$, $C=1$, $\lambda=0.5$ dan $\gamma=0.1$
- c. Menghitung matriks dengan rumus (2):

$$D_{ij} = y_i y_j (K(x_i \rightarrow \cdot \cdot x_j \rightarrow \cdot \cdot) + \lambda^2) \tag{2}$$

Keterangan :

- D_{ij} = elemen matriks data ke-ij
- y_i = label data ke-i
- y_j = label data ke-j
- $K(x_i \rightarrow \cdot \cdot x_j \rightarrow \cdot \cdot)$ = nilai kernel linear
- λ = lambda

- d. Untuk data ke $n = 1, 2, 3, \dots n$ gunakan persamaan (3),(4),(5) berikut.

$$E_i \sum_{j=1}^n \alpha_j D_{ij} \tag{3}$$

$$\delta \alpha_i = \min\{\max[\gamma(1 - E_i, -\alpha_i], C - \alpha_i\} \tag{4}$$

$$\alpha_i = \alpha_i + \delta \alpha_i \tag{5}$$

Keterangan :

- E = Nilai error
- α_j = alfa ke-j
- D_{ij} = matriks Hessian
- $\delta \alpha_i$ = delta alfa ke-i
- γ = learning rate
- α_i = alfa ke-i
- C = variable slack
- $\delta \alpha$ = delta alpha

- e. Mencari nilai bias (b) dengan menggunakan persamaan (6)

$$b = -\frac{1}{2} [w \cdot x^+ + w \cdot x^-] \tag{6}$$

- f. Pengujian pada dokumen uji
- g. Perhitungan Keputusan

$$h(x) = \begin{cases} +1 & \text{jika } w \cdot x + b \geq 0 \\ -1 & \text{jika } w \cdot x + b < 0 \end{cases} \tag{7}$$

Jika hasil perhitungan keputusan lebih dari sama dengan 0 maka nilai adalah +1, maka termasuk kelas positif dan jika hasil perhitungan keputusan kurang dari 0 maka nilai nilai -1, maka termasuk kelas negatif. Perhitungan keputusan dengan menggunakan persamaan (7).

$$h(x) = w \cdot x + b \tag{8}$$

2.2 Evaluasi

Fase evaluasi digunakan untuk menilai kebenaran pemodelan setelah diterapkan pada kumpulan data pelatihan. Menerapkan classification report untuk menentukan precision, recall, fi-score, dan accuracy akan memungkinkan Anda membandingkan hasil dari dua kumpulan data terpisah.

3. HASIL DAN PEMBAHASAN.

3.1 Persiapan Data

Penelitian ini mengambil data dari Twitter. Data yang diambil merupakan tweet pada rentang waktu 1 Januari 2023 sampai 13 Agustus 2023. Hasil dari proses scrapping akan diubah ke dalam ekstensi .csv untuk kemudian diolah menggunakan bahasa pemrograman python. Jumlah data yang diperoleh dari hasil scrapping adalah 1719



data. Dimana 597 data berasal dari kata kunci bakal calon 1 yaitu #aniesbaswedan, 627 data berasal dari bakal calon 2 yaitu #ganjarpranowo dan 495 data berasal dari bakal calon 3 yaitu #prabowosubianto.

3.2 Text Preprocessing

Persiapan teks dilakukan pada awalnya setelah pengumpulan data Twitter selesai. Dataset yang digunakan berisi data yang tidak terstruktur, oleh karena itu dilakukan hal ini. Daftar berikut menjelaskan berbagai langkah persiapan teks yang digunakan dalam penelitian ini:

- a. Case Folding adalah tahap pengubahan semua huruf campuran baik itu uppercase, atau lowercase, menjadi lowercase (huruf kecil) semua.
b. Filtering adalah tahap pembersihan data dari tanda baca, simbol atau elemen yang tidak dibutuhkan seperti URL.
c. Tokenizing adalah tahap pemecahan teks menjadi kata-kata, seperti "saya pergi ke sekolah" menjadi "saya", "pergi", "ke", "sekolah".
d. Stopword Removal adalah tahap penghapusan kata sambung seperti "ke", "di", "dan", "dia", "kami", "aku", "saya".

Pada tabel 1. dibawah ini disajikan contoh data uji setelah melewati tahap preprocessing sehingga diperoleh data uji yang siap dianalisis

Tabel 1. Hasil Text Preprocessing

Table with 2 columns: Input and Output. The input is a tweet about Indonesia's popularity, and the output is a list of extracted words like 'bilang', 'senang', 'orang', etc.

3.3 Vektorisasi TF-IDF

Metode vektorisasi memiliki keunggulan dalam memberikan nilai pada kata-kata dalam dokumen atau tweet, sehingga memungkinkan dilakukannya pemrosesan dan pencarian pola ketika dokumen atau tweet tersebut memasuki tahap klasifikasi. Dengan memeriksa terlebih dahulu sampel data tweet kemudian melakukan perhitungan berdasarkan rumus (1), metode vektorisasi menggunakan teknik perhitungan pembobotan kata atau TF-IDF. Berikut kesimpulan yang diperoleh setelah melakukan perhitungan TF-IDF terhadap data tweet dari ketiga sampel yang ditunjukkan pada Tabel 3.

Contoh :

Data Latih :

- D1 = ['dukung', 'presiden', 'pilih']
D2 = ['capres', 'bijak', 'lanjut', 'suka', 'rakyat']
D3 = ['pimpin', 'lanjut']
D4 = ['dungu', 'tolak', 'orang', 'langgar', 'abadi', 'rekam']
D5 = ['citra', 'blunder', 'rekam', 'pasang', 'presiden', 'blunder', 'indonesia']

Contoh hasil Vektorisasi :

Setelah tahap preprocessing langkah yang dilakukan selanjutnya adalah pembobotan TF-IDF. Adapun hasil pembobotan yang dilakukan dapat dilihat pada gambar 2. Berikut.

Tabel 2. Hasil Vektorisasi

Table with 6 columns: Term, D1, D2, D3, D4, D5. It shows TF-IDF values for various terms across five documents (D1-D5).

Term	TF-IDF				
	D1	D2	D3	D4	D5
rekam	0	0	0	0	0.206
pasang	0	0	0	0	0.206
blunder	0	0	0	0	0.412
indonesia	0	0	0	0	0.206

3.4 Proses Klasifikasi Support Vektor Machine

Untuk membangun suatu model yang selanjutnya akan digunakan untuk data uji, dilakukan proses klasifikasi dengan menggunakan data latih. Untuk setiap kumpulan data calon presiden, bagi data pelatihan menjadi 80% dan data uji menjadi 20% sebelum memulai prosedur klasifikasi.

Lima contoh tweet dengan variabel X sebagai istilah data masukan dan Y sebagai label ditunjukkan pada Tabel 3, di mana -1 menunjukkan label negatif dan 1 menunjukkan label positif. Algoritma SVM digunakan untuk menghitung data pelatihan. Sistem belajar dari hasil data pelatihan yang akan digunakan sebagai referensi ke depannya. Temuan data uji dapat digunakan untuk mengetahui apakah sebuah tweet termasuk dalam kategori baik atau negatif. Setiap data akan dibandingkan dengan data lain dan data itu sendiri selama tahap fungsi kernel; alhasil Tabel 3. menunjukkan jika datanya berjumlah 5.

Tabel 3. Fungsi Kernel

	A1	A2	A3	A4	A5
B1	K(A1,B1)	K(A2,B1)	K(A3,B1)	K(A4,B1)	K(A5,B1)
B2	K(A1,B2)	K(A2,B2)	K(A3,B2)	K(A4,B2)	K(A5,B2)
B3	K(A1,B3)	K(A2,B3)	K(A3,B3)	K(A4,B3)	K(A5,B3)
B4	K(A1,B4)	K(A2,B4)	K(A3,B4)	K(A4,B4)	K(A5,B4)
B5	K(A1,B5)	K(A2,B5)	K(A3,B5)	K(A4,B5)	K(A5,B5)

Contoh perhitungan untuk data pertama :

$$K(x,y) = x * y$$

$$K(x,y) = (t1d1 * t1d1 + t1d2 * t1d2 + t1d3 * t1d3 + t1d4 * t1d4 + t1d5 * t1d5)$$

$$K(x,y) = (0.206 * 0.206 + 0*0 + 0*0 + 0*0 + 0*0) = 0.042$$

Lakukan cara yang sama pada keseluruhan data sehingga hasilnya akan tampak seperti pada tabel 4. sebagai berikut.

Tabel 4. Perhitungan Kernel Linear

No.	1	2	3	..	18	19	20
1	0.042	0.034	0.042	..	0	0	0
2	0.034	0.055	0.034	..	0.034	0.068	0.034
3	0.042	0.034	0.042	..	0	0	0
4	0	0	0	.	0	0	0
5	0	0	0	..	0	0	0
6	0	0	0	..	0	0	0
7	0	0	0	..	0	0	0
8	0	0	0	..	0	0	0
9	0	0	0	..	0	0	0
10	0	0	0	..	0	0	0
11	0	0	0	..	0	0	0
12	0	0	0	..	0	0	0
13	0	0	0	..	0	0	0
14	0	0	0	..	0	0	0
15	0	0	0	..	0	0	0
16	0	0.034	0	..	0.042	0.085	0.042
17	0	0.034	0	..	0.042	0.085	0.042
18	0	0.034	0	..	0.042	0.085	0.042
19	0	0.068	0	..	0.085	0.169	0.085
20	0	0.034	0	..	0.042	0.085	0.042

Setelah nilai kernel diketahui, langkah selanjutnya adalah perhitungan matriks Hessian. Contoh perhitungan nilai matriks Hessian pada data pertama

$$\begin{aligned}
 D_{11} &= y_i y_j (K(x_i, x_j) + \lambda^2) \\
 &= 1 * 1(0.042) + 0.5^2 \\
 &= 0.292
 \end{aligned}$$



Lakukan cara yang sama pada keseluruhan data sehingga hasilnya akan tampak seperti pada tabel 5. sebagai berikut.

Tabel 5. Perhitungan Matriks Hessian

No.	1	2	3	..	18	19	20
1	0.292	0.284	0.292	..	0.25	0.25	0.25
2	0.284	0.305	0.284	..	0.284	0.318	0.284
3	0.292	0.284	0.292	..	0.25	0.25	0.25
4	0.25	0.25	0.25	..	0.25	0.25	0.25
5	0.25	0.25	0.25	..	0.25	0.25	0.25
6	0.25	0.25	0.25	..	0.25	0.25	0.25
7	0.25	0.25	0.25	..	0.25	0.25	0.25
8	0.25	0.25	0.25	..	0.25	0.25	0.25
9	0.25	0.25	0.25	..	0.25	0.25	0.25
10	-0.25	-0.25	-0.25	..	-0.25	-0.25	-0.25
11	0.25	0.25	0.25	..	0.25	0.25	0.25
12	0.25	0.25	0.25	..	0.25	0.25	0.25
13	0.25	0.25	0.25	..	0.25	0.25	0.25
14	-0.25	-0.25	-0.25	..	-0.25	-0.25	-0.25
15	-0.25	-0.25	-0.25	..	-0.25	-0.25	-0.25
16	0.25	0.284	0.25	..	0.292	0.335	0.292
17	0.25	0.284	0.25	..	0.292	0.335	0.292
18	0.25	0.284	0.25	..	0.292	0.335	0.292
19	0.25	0.318	0.25	..	0.335	0.419	0.335
20	0.25	0.284	0.25	..	0.292	0.335	0.292

Setelah kita mendapatkan nilai matriks Hessian, kita dapat menggunakan fungsi substitusi matriks persamaan (3) untuk mencari nilai errornya. Hasilnya dapat dilihat pada tabel 6. berikut.

Tabel 6. Perhitungan Nilai Error

No.	E_i
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0

Temuannya ditunjukkan pada Tabel 7. di bawah ini. Selanjutnya gunakan persamaan (4) untuk mendapatkan nilai delta alpha.

Tabel 7. Perhitungan Nilai Delta Alpha

No.	$\delta\alpha$
1	0.1
2	0.1
3	0.1
4	0.1
5	0.1



No.	$\delta\alpha$
6	0.1
7	0.1
8	0.1
9	0.1
10	0.1
11	0.1
12	0.1
13	0.1
14	0.1
15	0.1
16	0.1
17	0.1
18	0.1
19	0.1
20	0.1

Temuan pada Tabel 8. kemudian diperoleh dengan menerapkan persamaan (5) untuk menghitung nilai alpha baru:

Tabel 8. Perhitungan Nilai Alpha Baru

No.	α
1	0.1
2	0.1
3	0.1
4	0.1
5	0.1
6	0.1
7	0.1
8	0.1
9	0.1
10	0.1
11	0.1
12	0.1
13	0.1
14	0.1
15	0.1
16	0.1
17	0.1
18	0.1
19	0.1
20	0.1

Lakukan proses pencarian nilai alpha baru dengan mengulang proses pencarian nilai error, ulangi proses sampai dengan batas iterasi yang ditetapkan, pada contoh ini ditetapkan 2 kali iterasi maksimal, sehingga pada iterasi ke-2 akan diperoleh nilai alpha baru seperti pada tabel 9. berikut ini.

Tabel 9. Nilai Alpha pada Iterasi maksimal

No.	α
1	0.257893
2	0.255541
3	0.257893
4	0.256946
5	0.256946
6	0.256688
7	0.256946
8	0.256946
9	0.258367
10	0.343225
11	0.259225
12	0.259225
13	0.259225



No.	α
14	0.343225
15	0.343225
16	0.25599
17	0.25599
18	0.25599
19	0.252738
20	0.25599

Nilai bias akan ditentukan menggunakan persamaan (6) pada langkah berikut, namun terlebih dahulu harus menghitung nilai w . Perhitungannya adalah sebagai berikut, dimana w_1+ mewakili bobot data perkalian titik pada kelas positif dan w_1- mewakili bobot data perkalian titik pada kelas negatif.

$$w \cdot x^+ (\text{kelas positif}) = (0.2579 * 1 * 0.25) + (0.2555 * 1 * 0.25) + (0.2579 * 1 * 0.25) + (0.25695 * 1 * 0.25) + (0.25695 * 1 * 0.25) + (0.25669 * 1 * 0.25) + (0.25695 * 1 * 0.25) + (0.25695 * 1 * 0.25) + (0.25837 * 1 * 0.25) + (0.34323 * -1 * -0.292) + (0.25923 * 1 * 0.292) + (0.25923 * 1 * 0.292) + (0.25923 * 1 * 0.292) + (0.34323 * -1 * -0.292) + (0.34323 * -1 * -0.292) + (0.25599 * 1 * 0.25) + (0.25599 * 1 * 0.25) + (0.25599 * 1 * 0.25) + (0.25274 * 1 * 0.25) + (0.25599 * 1 * 0.25) = 1.425463$$

$$w \cdot x^- (\text{kelas negatif}) = (0.2579 * 1 * -0.25) + (0.2555 * 1 * -0.25) + (0.2579 * 1 * -0.25) + (0.25695 * 1 * -0.25) + (0.25695 * 1 * -0.25) + (0.25669 * 1 * -0.25) + (0.25695 * 1 * -0.25) + (0.25695 * 1 * -0.25) + (0.25837 * 1 * -0.25) + (0.34323 * -1 * 0.292) + (0.25923 * 1 * -0.292) + (0.25923 * 1 * -0.292) + (0.25923 * 1 * -0.292) + (0.34323 * -1 * 0.292) + (0.34323 * -1 * 0.292) + (0.25599 * 1 * -0.25) + (0.25599 * 1 * -0.25) + (0.25599 * 1 * -0.25) + (0.25274 * 1 * -0.25) + (0.25599 * 1 * -0.25) = -1.42546$$

Jika nilai $w \cdot x^+$ (kelas positif) dan $w \cdot x^-$ (kelas negatif) sudah diketahui, maka nilai b dapat dicari dengan rumus persamaan (6):

$$b = -1/2 (w \cdot x^+ + w \cdot x^-) = 0.000015$$

Setelah nilai α , Anda dapat menguji sampel data uji di bawah ini.

D1 = ['lebih', 'baik', 'capres', 'pimpin', 'lanjut', 'bijak', 'pimpin', 'auto', 'coblos', 'capres']

Dengan menggunakan persamaan 1 untuk mencari nilai TF-IDF, maka diperoleh hasil TF-IDF dari setiap term pada data uji seperti yang dapat dilihat pada tabel 10. berikut.

Tabel 10. TF-IDF Data Uji

TF-IDF	1	2	3	4	5	6	7	8
D1	0.2673	0.2673	0.4019	0.4019	0.3243	0.2673	0.2673	0.2673

Langkah selanjutnya adalah mengalikan data uji dengan seluruh data pelatihan untuk menghasilkan perkalian titik, yang kemudian dikirim ke fungsi kernel. Sehingga akan diperoleh nilai dari dot product data uji dengan data latih seperti pada tabel 11. Dibawah ini.

Tabel 11. Dot product data uji dengan data latih

No.	Dot Product
1	0.213
2	0.213
3	0.32
4	0.32
5	0.259
6	0.213
7	0.213
8	0.213

Setelah mendapatkan dot product antara data uji ,barulah dilakukan perhitungan fungsi keputusan dengan menggunakan rumus (7).

$$f(x) = w \cdot x + b = \sum \alpha_i y_i K(x_i, x_j) + b = (0.213 + 0.213 + 0.32 + 0.32 + 0.259 + 0.213 + 0.32 + 0.213 + 0.213 + 0.32) + (- 0.000015) = 2.603$$

Fungsi klasifikasi = sign (2.603) = +1

Setelah melakukan proses pengujian pada data uji, diketahui bahwa pada fungsi klasifikasi data mendapatkan nilai positif 1 sehingga di klasifikasikan sebagai kategori kelas 1 dimana kelas 1 adalah kelas **Positif**.

3.5 Evaluasi

Setelah melakukan pemodelan dengan metode SVM pada seluruh dataset, yaitu dataset dari sumber #aniesbaswedan, #ganjarpranowo dan #prabowosubianto, maka diperoleh classification report seperti pada gambar dibawah ini. Dimana hasil dari penelitian berupa accuracy, precision, recall dan f1-score akan disajikan dalam bentuk classification report.

Pada analisis sentimen yang dilakukan untuk calon 1 dengan menggunakan hashtag #aniesbaswedan maka diperoleh hasil analisis seperti pada classification report dibawah ini. Dimana dapat dilihat bahwa nilai akurasi algoritma Support Vector Machine pada studi kasus ini dengan menggunakan rasio data latih dan data uji 7:8 adalah sebesar 75% seperti yang dapat dilihat pada gambar dibawah 3. dibawah.

	precision	recall	f1-score	support
negatif	0.76	0.86	0.81	73
positif	0.71	0.56	0.63	45
accuracy			0.75	118
macro avg	0.74	0.71	0.72	118
weighted avg	0.74	0.75	0.74	118

Gambar 3. Classification report pengujian bakal calon 1

Sedangkan untuk sentimen yang ditujukan kepada calon 2 yaitu hashtag #ganjarpranowo, proses analisis yang dilakukan memperoleh nilai akurasi yang lebih besar. Nilai akurasi yang diperoleh untuk sentimen pada calon 2 adalah 86% dapat dilihat pada gambar 4.

	precision	recall	f1-score	support
negatif	1.00	0.06	0.11	18
positif	0.86	1.00	0.93	107
accuracy			0.86	125
macro avg	0.93	0.53	0.52	125
weighted avg	0.88	0.86	0.81	125

Gambar 4. Classification report pengujian bakal calon 2

Untuk sentimen pada calon 3 yaitu dari hashtag #prabowosubianto, diperoleh nilai akurasi sebesar 72%. Adapun hasil classification report dapat dilihat pada gambar 5. berikut ini.

	precision	recall	f1-score	support
negatif	0.68	0.51	0.58	37
positif	0.74	0.85	0.79	60
accuracy			0.72	97
macro avg	0.71	0.68	0.69	97
weighted avg	0.72	0.72	0.71	97

Gambar 5. Classification report pengujian bakal calon 3

4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah diuraikan di atas, proses analisis sentimen terhadap sentimen twitter mengenai bakal calon presiden republik Indonesia tahun 2024 dapat dilakukan dengan metode Support Vector Machine. Total sentimen awal yang dianalisis berjumlah 1719 buah sentimen. Dengan persebaran 597 data untuk aniesbaswedan, 627 data untuk ganjarpranowo dan 495 data untuk prabowo subianto. Proses analisis yang dilakukan memiliki nilai akurasi 75% pada sentimen calon 1, 86% pada calon 2 dan 72% pada calon 3. Sehingga rata-rata akurasi untuk ketiga calon adalah sebesar 78%. Hasil akurasi terhadap penerapan metode support vector machine dipengaruhi oleh beberapa hal, salah satunya adalah jumlah dataset dan komposisi jumlah sentimen positif dan negatif

UCAPAN TERIMAKASIH

Saya ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dalam penelitian ini. Terimakasih kepada keluarga saya atas dukungan moral dan doa-doanya yang tak ternilai. Terima kasih pula kepada dosen pembimbing yang telah memberikan bimbingan, nasihat, dan waktunya selama proses penelitian ini. Saya juga ingin mengucapkan terima kasih kepada teman-teman sejawat yang telah



berbagi pengetahuan dan pengalaman mereka, serta semua pihak yang telah memberikan akses dan bantuan dalam pengumpulan data. Penelitian ini tidak mungkin terlaksana tanpa dukungan dari semua pihak ini. Semoga hasil penelitian ini dapat bermanfaat bagi semua. Sekali lagi, terima kasih banyak atas segala dukungan yang telah diberikan

REFERENCES

- [1] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.
- [2] M. R. Fahlevi, "Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi Dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia Di Google Playstore Menggunakan Metode Support Vector Machine," *J. Teknol. dan Komun. Pemerintah.*, vol. 4, no. 1, pp. 1–13, 2022, doi: 10.33701/jtkp.v4i1.2701.
- [3] G. T. Santos, "Analisis Sentimen Pada Tweet Dengan Tagar #Bpjsrasarentenir Menggunakan Metode Support Vectore Machine (Svm)," pp. 12–13, 2021.
- [4] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
- [5] S. Nurul, J. Fitriyyah, N. Safrjadi, 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," *(Jurnal Edukasi dan Penelit. Inform.*, vol. 5, no. 3, pp. 279–285, 2019, [Online]. Available: <http://dx.doi.org/10.26418/jp.v5i3.34368>.
- [6] H. Setiawan, E. Utami, and S. Sudarmawan, "Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes," *J. Komtika (Komputasi dan Inform.*, vol. 5, no. 1, pp. 43–51, 2021, doi: 10.31603/komtika.v5i1.5189.
- [7] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.
- [8] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *J. Inform. Upgris*, vol. 7, no. 1, pp. 36–40, 2021.
- [9] A. S. Rahayu, A. Fauzi, and R. Rahmat, "Komparasi Algoritma Naive Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 349, 2022, doi: 10.30865/json.v4i2.5398.
- [10] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naive Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [11] T. Berdasarkan and M. Metode, "Analisis Sentimen Pemilihan Presiden Indonesia Tahun 2019 Di Twitter Berdasarkan Geolocation Menggunakan Metode Naive Bayesian Classification Sentiment Analysis of Indonesia Presidency Election 2019 on Twitter Based on Geolocation Using Naive Bayesian Cla," pp. 115–127, 2019.
- [12] M. Raihan, F. Sya, U. Enri, and T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma Naive Bayes," vol. 9, no. 2, pp. 265–273, 2022, doi: 10.30865/jurikom.v9i2.3989.
- [13] S. F. Pratama, R. Andrian, and A. Nugroho, "Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 4, no. 2, p. 39, 2019, doi: 10.31328/jointecs.v4i2.1004.
- [14] W. P. Jati, A. Z. Lutfi, H. Ilma, and E. Widodo, "Analisis Sentimen terhadap Isu Politik Calon Presiden Indonesia 2019 berdasarkan Opini Netizen dari Twitter menggunakan Metode Klasifikasi Support Vector Machine," *Pros. Konf. Nas. Penelit. Mat. dan Pembelajarannya IV 2019*, no. Prodi Pendidikan Matematika FKIP UMS, p. PROSIDING-M2, 2019.
- [15] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine," *Komputek*, vol. 3, no. 2, p. 52, 2019, doi: 10.24269/jkt.v3i2.270.
- [16] M. Furqan, S. Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.Com*, vol. 21, no. 1, pp. 51–60, 2022, doi: 10.33633/tc.v21i1.5446.
- [17] M. S. Hasibuan and A. Serdano, "Analisis Sentimen Kebijakan Pembelajaran Tatap Muka Menggunakan Support Vector Machine dan Naive Bayes," *JRST (Jurnal Ris. Sains dan Teknol.*, vol. 6, no. 2, p. 199, 2022, doi: 10.30595/jrst.v6i2.15145.
- [18] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.
- [19] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, T. Tukino, and B. Priyatna, "Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 8, no. 1, p. 11, 2023, doi: 10.31328/jointecs.v8i1.4265.
- [20] H. Sujadi, "Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine," *INFOTECH J.*, vol. 8, no. 1, pp. 22–27, 2022, doi: 10.31949/infotech.v8i1.1883.