

Public Sentiment Analysis on the Launch of Danantara Using Support Vector Machine (SVM) Algorithm

Ziagy Aji Fernando ¹, Rumini ², Tri Susanto ³

^{1,2,3} Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Yogyakarta, 55283, Indonesia

Informasi Artikel

Diterima : 16 Desember 2025
Revisi : 25 Desember 2025
Publikasi : 31 Desember 2025

Kata Kunci:

Support Vector Machine (SVM)
Danantara
Sentimen Publik
Media Sosial X

ABSTRAK

Pengelolaan kekayaan dan sumber daya negara secara efisien menjadi kunci dalam mewujudkan kesejahteraan rakyat Indonesia. Badan Pengelola Investasi Daya Anagata Nusantara (BPI Danantara) hadir sebagai lembaga strategis yang bertujuan mengoptimalkan investasi pemerintah untuk mendorong pertumbuhan ekonomi nasional. Peluncuran Proyek Danantara memicu berbagai tanggapan publik, khususnya di media sosial X, yang kini menjadi ruang ekspresi opini masyarakat terhadap kebijakan pemerintah. Penelitian ini bertujuan menganalisis sentimen publik terhadap Proyek Danantara menggunakan algoritma *Support Vector Machine (SVM)*. Dengan mengolah 3.152 data postingan dari periode 24 Februari hingga 4 April 2025, model SVM menunjukkan performa tinggi dengan akurasi 87%. Hasil penelitian menunjukkan bahwa sentimen positif lebih dominan secara kuantitatif, mencerminkan ekspektasi publik terhadap kemajuan investasi nasional dan pengelolaan strategis proyek. Namun, sentimen negatif juga cukup signifikan, menyoroti isu korupsi, ketidaktransparanan, serta potensi kegagalan proyek.

ABSTRACT

Efficient management of national wealth and resources is key to realizing the welfare of the Indonesian people. The Daya Anagata Nusantara Investment Management Agency (BPI Danantara) was established as a strategic institution aimed at optimizing government investments to drive national economic growth. The launch of the Danantara Project has sparked various public responses, particularly on the social media platform X, which has become a space for citizens to express their opinions on government policies. This study aims to analyze public sentiment toward the Danantara Project using the Support Vector Machine (SVM) algorithm. By processing 3,152 posts from the period of February 24 to April 4, 2025, the SVM model demonstrated strong performance with an accuracy of 87%. The research findings indicate that positive sentiment is quantitatively more dominant, reflecting public expectations of national investment progress and strategic project management. However, negative sentiment is also quite significant, highlighting issues of corruption, lack of transparency, and potential project failure.

This is an open-access article under the [CC BY-SA](#) license



*Penulis Koresponden

Email: rumini@amikom.ac.id

Cara sitasi IEEE::

Z. A. Fernando, R. Rumini, T. Susanto, "Public Sentiment Analysis on the Launch of Danantara Using Support Vector Machine (SVM) Algorithm," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 4, p. 1425-11433, Desember 2025. doi: 10.30811/jaise.v5i4.8496

1. PENDAHULUAN

Pengelolaan kekayaan dan sumber daya negara secara efisien merupakan fondasi penting dalam upaya mewujudkan kesejahteraan yang merata bagi seluruh rakyat Indonesia. Dalam konteks tersebut, pemerintah membentuk Badan Pengelola Investasi Daya Anagata Nusantara (BPI Danantara) sebagai lembaga strategis yang bertugas mengelola dan mengoptimalkan investasi nasional. Nama "Daya Anagata Nusantara" yang diberikan oleh Presiden Prabowo Subianto mengandung makna simbolik yang mencerminkan energi, masa depan, dan kedaulatan bangsa. Melalui tata kelola yang baik, efisiensi aset, dan peningkatan daya saing, Danantara Indonesia berkomitmen mendorong transformasi ekonomi serta menarik investasi global demi tercapainya kemakmuran yang inklusif [1].

Peluncuran Proyek Danantara menjadi perhatian publik, terutama di media sosial X yang kini berfungsi sebagai ruang diskusi terbuka terhadap kebijakan pemerintah [2]. Fenomena ini serupa dengan respons masyarakat terhadap program makan siang gratis yang sempat menjadi perbincangan hangat selama masa pemilu. Reaksi publik terhadap Proyek Danantara pun beragam, mulai dari dukungan hingga kritik. Untuk memahami persepsi masyarakat secara lebih mendalam, pendekatan analisis sentimen menjadi metode yang relevan dan efektif [3].

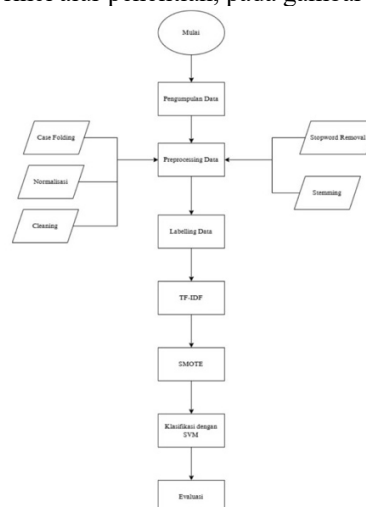
Analisis Sentimen adalah sebuah metode dalam bidang pemrosesan Bahasa alami (NLP) yang bertujuan untuk mengenali dan mengelompokkan sikap atau perasaan individu mengenai suatu hal melalui data teks. Belakangan ini, penerapannya semakin meluas di berbagai sektor, seperti mengevaluasi tanggapan masyarakat terhadap kebijakan pemerintah, menilai ulasan produk, hingga menganalisis persepsi terhadap masalah-masalah sosial.

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap Proyek Danantara dengan memanfaatkan algoritma Support Vector Machine (SVM). Algoritma ini dipilih karena kemampuannya dalam mengklasifikasikan teks secara akurat dan telah terbukti efektif dalam berbagai studi sebelumnya, dengan tingkat akurasi yang mencapai hingga 98% [4]. Berdasarkan penelitian yang dilakukan oleh Shahmirul Hafizullah Imanuddin. (2023) menunjukkan bahwa Analisis sentimen menggunakan algoritma Support Vector Machine (SVM) mengklasifikasikan ulasan menjadi sentimen. Hasil penelitian menunjukkan SVM mampu mencapai akurasi sebesar 91%. Temuan ini menguatkan bahwa SVM merupakan alat yang efektif dalam mengidentifikasi pola sentimen pengguna[7]. Temuan ini semakin dikuatkan oleh penelitian Prabowo dan Azizah (2020) yang menerapkan metode Support Vector Machine (SVM) untuk mengidentifikasi ujaran kebencian dan *cyberbullying* di platform media sosial, sekaligus menunjukkan keberhasilan dalam klasifikasi sentimen[8].

Beberapa penelitian juga menunjukkan bahwa SVM mampu mengelompokkan opini publik secara sistematis dan memberikan wawasan yang berguna dalam memahami argumen masyarakat [5][6]. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam evaluasi kebijakan publik serta memperkuat penerapan *machine learning* dalam analisis sosial-politik berbasis media digital.

2. METODE

Bagian metode memaparkan secara rinci alur penelitian, pada gambar 1.



Gambar 1. Alur Penelitian

Pada Gambar 1 Alur penelitian, dapat dijelaskan sebagai berikut :

2.1 Dataset

Data yang digunakan dalam penelitian ini bersifat data primer, yang diperoleh langsung dari unggahan pengguna di platform X yang membahas proyek Danantara. Pengambilan data dilakukan dengan teknik crawling menggunakan library Python dan menggunakan kode yang dibangun dengan Node.js berdasarkan kata kunci “Danantara id” dan “Daya Anagata Nusantara”. Dataset berasal dari komentar unggahan (tweet) yang menyebutkan kata kunci terkait proyek Danantara pada platform X sejak waktu peresmiannya pada 24 Februari 2025 sampai dengan 5 Mei 2025. Sementara itu, sampel yang digunakan terdiri dari 2066 unggahan yang berhasil dikumpulkan.

2.2 Pre-Processing Data

Tahap preprocessing data merupakan proses pengolahan data mentah sebelum dilakukan pembobotan dan analisis lebih lanjut. Preprocessing bertujuan untuk membersihkan dan menyederhanakan data teks agar lebih mudah diproses oleh algoritma machine learning. Adapun tahapan-tahapan preprocessing yang dilakukan dalam penelitian ini meliputi:

1. *Case Folding*
Mengubah seluruh huruf dalam tweet menjadi huruf kecil (lowercase) untuk menyamakan bentuk kata dan menghindari perbedaan akibat kapitalisasi.
2. *Cleaning*
Menghapus karakter atau elemen yang tidak dibutuhkan seperti tanda baca, angka, emoji, URL, username (@user), hashtag (#), dan karakter khusus lainnya.
3. *Normalisasi*
Proses mengubah data mentah ke dalam bentuk yang lebih seragam, bersih, dan konsisten, agar mudah dianalisis atau diproses oleh model.
4. *Stopword Removal*
Menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti “yang”, “dan”, “di”, dan sebagainya.
5. *Stemming*
Mengubah kata-kata menjadi bentuk dasarnya (root word), sehingga kata “membayar”, “pembayaran”, dan “dibayar” akan dikembalikan ke bentuk dasar “bayar”.

2.3 Pelabelan Data

Pelabelan data merupakan proses pemberian label sentimen pada setiap tweet yang dikumpulkan, dengan tujuan untuk mengklasifikasikan opini pengguna terhadap proyek Danantara. Dalam penelitian ini, pelabelan dilakukan secara otomatis menggunakan bantuan library Lexicon. Pelabelan dikelompokkan menjadi 2 kelas yaitu positif dan negatif. Proses pelabelan dilakukan dengan menjumlahkan skor sentimen dari setiap kata dalam tweet. Skor akhir kemudian digunakan untuk menentukan kelas sentimen:

1. Jika skor > 0 maka tweet diberi label positif
2. Jika skor < 0 maka tweet diberi label negatif

2.4 Pembobotan Kata (TF-IDF)

Pembobotan kata dalam penelitian ini dilakukan dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini untuk memberikan nilai numerik pada setiap kata dalam tweet berdasarkan seberapa sering kata tersebut muncul dalam satu dokumen (*term frequency*) dan seberapa jarang kata tersebut muncul di seluruh kumpulan dokumen (*inverse document frequency*).

TF-IDF tidak hanya meningkatkan akurasi pengindeksan dan pencarian dokumen dengan menyoroti kata kunci yang relevan, tetapi juga memperkaya analisis sentimen dengan mengidentifikasi istilah yang berpengaruh kuat dalam mempresentasikan opini atau emosi teks[22][23]. Rumus metode Term Frequency-Inverse Document Frequency (TF-IDF) pada rumus (1):

$$\begin{aligned}
 tf &= 0,5 + 0,5 \times \frac{tf}{\max(tf)} \\
 idf_t &= \log\left(\frac{D}{df_t}\right) \\
 W_{d,t} &= tf_{d,t} \times idf_{d,t}
 \end{aligned}
 \tag{1}$$

Keterangan :

- D = dokumen ke-d
- t = *term* ke-t dari dokumen
- W = bobot ke-d terhadap *term* ke-t
- tf = jumlah kemunculan *term* i dalam dokumen
- idf = *Inversed Document Frequency*
- df = banyak dokumen yang mengandung term i

2.5 SMOTE

Dalam penelitian ini, distribusi data sentimen yang diperoleh tidak seimbang, dimana jumlah data pada satu kelas lebih sedikit dibandingkan dua kelas lainnya. Ketidakeimbangan data ini dapat mempengaruhi kinerja model klasifikasi dan menyebabkan model lebih cenderung memprediksi kelas mayoritas. Untuk mengatasi masalah ini, digunakan metode SMOTE (*Synthetic Minority Over-Sampling Tehnique*). SMOTE bekerja dengan cara membuat data sintetis (buatan) untuk kelas minoritas, berdasarkan kemiripan data dalam ruang fitur.

2.6 Klasifikasi dengan Metode *Support vector Machine* (SVM)

Pembuatan model menggunakan algoritma *Support Vector Machine* (SVM). SVM merupakan algoritma supervised learning yang digunakan untuk menyelesaikan permasalahan klasifikasi maupun regresi. Dalam konteks klasifikasi sentimen, SVM bekerja dengan mencari hyperplane terbaik yang dapat memisahkan data ke dalam beberapa kelas secara optimal, dalam hal ini kelas positif, dan negatif.

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang efektif untuk tugas klasifikasi dan regresi. Metode ini bekerja dengan memaksimalkan jarak antara kelas-kelas data dalam ruang fitur, sehingga mengurangi risiko kesalahan prediksi pada data uji yang bel pernah dilihat sebelumnya. Berdasarkan teori pembelajaran statistic, SVM membangun model prediktif yang optimal dengan menemukan hyperplane terbaik yang memisahkan data secara jelas. Keunggulan ini membuat SVM banyak digunakan dalam scenario kompleks seperti pengenalan pola, analisis citra, atau klasifikasi teks[24].

Algoritma SVM juga menggunakan kernel yang mentransformasi data ke ruang fitur berdimensi lebih tinggi. Jenis-jenis kernel umum termasuk kernel linear, kernel polinomial, dan kernel radial basis function (RBF), masing-masing dengan parameter yang mempengaruhi kinerja dan fleksibilitas pada model secara keseluruhan. Kernel linear cocok untuk kasus sederhana karena menghitung dot produk antar fitur dan efektif untuk data yang dapat dipisahkan secara linear, sementara kernel polinomial mampu menangani batas keputusan nonlinear dengan bentuk fitur-fitur polinomial. Serta kernel yang paling populer yaitu kernel RBF dapat mengukur jarak antar dua titik dan sangat efektif untuk menangani data dengan hubungan kompleks dan tidak linear di ruang fitur[25].

Rumus persamaan *HyperPlane* pada klasifikasi SVM (2):

$$w \cdot x + b = 0 \tag{2}$$

Keterangan :

- w = vektor bobot (weight)
- x = vektor weight fitur input
- b = bias (intersep dari hyperplane)

Dalam konteks klasifikasi sentimen, SVM bekerja dengan mencari hyperplane terbaik yang dapat memisahkan data ke dalam beberapa kelas secara optimal, dalam hal ini kelas positif, dan negatif.

Contoh perhitungan kelas positif secara manual menggunakan algoritma SVM: Kalimat sentimen “Sikap positif Prabowo di Danantara bener-bener jadi pembeda dengan proyek lain”:

$$= (0.5 \cdot 0.15) + (1.8 \cdot 0.2) + (1.2 \cdot 0.2) + (0.9 \cdot 0.2) + (0.6 \cdot 0.15) + (-0.4 \cdot 0.15) + (0.2 \cdot 0.1) + 0.3$$

$$= 0.075 + 0.36 + 0.24 + 0.18 + 0.09 - 0.06 + 0.02 + 0.3 = 1.205 \text{ (Sentimen Positif)}$$

Kalimat sentiment “korupsi proyek - proyek yang urgensinya rendah kaya IKN membangun Danantara”:

$$= (-1.5 \cdot 0.3) + (-0.8 \cdot 0.25) + (-0.8 \cdot 0.25) + (-0.5 \cdot 0.2) + (-0.6 \cdot 0.2) + (-0.9 \cdot 0.15) + (0.7 \cdot 0.2) + (0.8 \cdot 0.15) + 0.3$$

$$= -0.45 - 0.2 - 0.1 - 0.12 - 0.135 + 0.14 + 0.12 + 0.3 = -0.445 \text{ (Sentimen Negatif)}$$

Dapat dilihat pada perhitungan manual tersebut kalimat “Sikap positif Prabowo di Danantara bener-bener jadi pembeda dengan proyek lain” dan kalimat “korupsi proyek - proyek yang urgensinya rendah kaya IKN membangun Danantara” dianalisis dengan pendekatan hyperlane. Pada metode ini, setiap kata penting dalam kalimat dipresentasikan sebagai fitur dengan nilai TF-IDF, kemudian dikalikan dengan bobot masing-masing fitur yang diperoleh dari hasil pelatihan SVM.

Nilai akhir dapat menunjukkan bahwa hasil klasifikasi model SVM berdasarkan nilai lebih besar atau lebih kurang dari 0. Dimana jika hasilnya lebih dari 0 akan dikategorikan sebagai sentimen positif dan jika hasilnya dibawah 0 maka akan dikategorikan sebagai sentimen negatif.

2.7 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui sejauh mana performa model klasifikasi dalam mengklasifikasikan data sentimen secara akurat. Pada penelitian ini, evaluasi dilakukan menggunakan metode *Confusion Matrix*. *Confusion Matrix* merupakan teknik evaluasi yang menyajikan perbandingan antara hasil prediksi model dengan label sebenarnya dalam bentuk matriks. Evaluasi ini penting untuk mengetahui performa model dan memastikan bahwa model mampu melakukan klasifikasi dengan baik, terutama pada data yang memiliki distribusi kelas yang tidak seimbang.

3. HASIL DAN PEMBAHASAN

Dari alur penelitian yang telah dilakukan, berikut ini adalah hasil penelitian yang telah dilakukan dan pembahasan.

3.1. Dataset

Penelitian ini menggunakan data unggahan pengguna di platform X yang berkaitan dengan proyek Danantara. Data dikumpulkan melalui proses crawling dengan bantuan tool tweet-harvest. Proses ini dilakukan dengan menggunakan bahasa pemrograman Phyton yang berjalan di lingkungan terminal atau notebook interaktif. Sumber data berasal dari tweet publik yang memuat kata kunci proyek Danantara, dibatasi oleh rentang waktu 24 Februari 2025 saat Danantara diluncurkan sampai 5 Mei 2025. Data yang diperoleh sebanyak 2066 kemudian disimpan dalam format CSV. Kode crawling data dan hasilnya dapat dilihat pada Gambar 2 dan Gambar 3.

```
# Crawl Data
filename = 'Danantaraid_raw.csv'
search_keyword = 'proyek danantara since:2025-01-01 until:2025-05-05 lang:id'
limit = 3000

Inpx -y tweet-harvest@2.6.1 -o '{filename}' -s '{search_keyword}' --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Gambar 2. Source Code Tweet-Harvest

	full_text	username	created_at	text_length	text_case_folding	full_text_bersih
0	Dengan dukungan dana 4 miliar USD dari QIA Dan...	Kkatinase	2025-05-04 06:51:07+00:00	206	dengan dukungan dana 4 miliar usd dari qia dan...	dengan dukungan dana miliar usd dari qia danan...
1	Danantara akan bangun Mega Proyek hasil etika...	WaneyFoto	2025-05-03 22:28:59+00:00	83	danantara akan bangun mega proyek hasil etika...	danantara akan bangun mega proyek hasil etika...
2	Sri Danantara tengah menjajaki peluang investa...	WaneyFoto	2025-05-03 23:13:57+00:00	144	bpi danantara tengah menjajaki peluang investa...	bpi danantara tengah menjajaki peluang investa...
3	@ARSIPAJA Buat danantara bisa cepa? Itu ngump...	mamatgcn_Z	2025-05-03 08:21:15+00:00	224	@arsipaja buat danantara bisa cepa? Itu ngump...	buat danantara bisa cepa? itu ngumpuln dana r...
4	Danantara bekerja sama dengan Qatar Investment...	ATLSERVICESUJ	2025-05-03 05:45:19+00:00	303	danantara bekerja sama dengan qatar investment...	danantara bekerja sama dengan qatar investment...

Gambar 3. Hasil Pengambilan Data

3.2. Pre-Processing Data

Tahap preprocessing data pada penelitian ini bertujuan untuk membersihkan dan mempersiapkan data mentah dari hasil crawling agar dapat diproses secara optimal oleh model machine learning. Tweet yang diperoleh dari X umumnya memiliki karakteristik yang tidak terstruktur, seperti penggunaan simbol, tautan, mention, hingga tanda baca berlebih yang dapat mengganggu analisis teks, sehingga dilakukan beberapa tahapan preprocessing yang terdiri dari: *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*.

3.2.1 Case Folding dan Cleaning

Pada tahap ini, semua huruf dalam teks diubah menjadi huruf kecil (lowercase). Tujuannya adalah untuk menyamakan format penulisan agar tidak terjadi perbedaan pemrosesan akibat huruf besar dan kecil. Setelah itu dilakukan pembersihan teks dari berbagai elemen yang tidak relevan agar data menjadi lebih bersih dan siap untuk dianalisis. Tujuan dari prosesi ini adalah menghasilkan teks yang lebih bersih dan konsisten pada gambar 4.

	full_text	full_text_bersih
0	Dengan dukungan dana 4 miliar USD dari QIA Dan...	dengan dukungan dana miliar usd dari qia danan...
1	Danantara akan bangun Mega Proyek hasil efisie...	danantara akan bangun mega proyek hasil efisie...
2	BPI Danantara tengah menjajaki peluang investa...	bpi danantara tengah menjajaki peluang investa...
3	@ARSIPAJA Buat danantara bisa cepat? Itu ngump...	buat danantara bisa cepat itu ngumpulin dana r...
4	Danantara bekerja sama dengan Qatar Investment...	danantara bekerja sama dengan qatar investment...

Gambar 4. Hasil Case Folding & Cleaning

3.2.2 Normalisasi

Normalisasi merupakan proses mengubah kata tidak baku, singkatan, slang menjadi bentuk standar atau baku. Pada tahap ini, data teks menjadi seragam, yang kemudian akan digunakan untuk analisis selanjutnya. Hasil proses Normalisasi ditunjukkan pada Gambar 5.

	full_text	text_normalisasi
0	Dengan dukungan dana 4 miliar USD dari QIA Dan...	dengan dukungan dana miliar usd dari qia danan...
1	Danantara akan bangun Mega Proyek hasil efisie...	danantara akan bangun mega proyek hasil efisie...
2	BPI Danantara tengah menjajaki peluang investa...	bpi danantara tengah menjajaki peluang investa...
3	@ARSIPAJA Buat danantara bisa cepat? Itu ngump...	buat danantara bisa cepat itu ngumpulin dana r...
4	Danantara bekerja sama dengan Qatar Investment...	danantara bekerja sama dengan qatar investment...
5	@Not_Ur_Bro @chaatimeeee @bembengkedua @jinjam...	sesakti yang bagaimana sih menurut parameter l...
6	@tempodotco Urusan sekolah layak 30 thn? Tp yg...	urusan sekolah layak tahun tapi yang ini knapa...

Gambar 5. Hasil Normalisasi

3.2.3 Stopword Removal dan Stemming

Tahap selanjutnya adalah *stopword Removal*. Proses ini bertujuan untuk menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis sentiment. Sehingga, dapat mengurangi dimensi data dan fokus pada kata-kata yang memiliki makna kuat terhadap opini pengguna dalam konteks proyek Danantara. Setelah dilakukan *Stopword Removal* tahap selanjutnya adalah *stemming* yang berfungsi untuk mengubah kata menjadi bentuk dasarnya. Hasil dari *stopword* dan *stemming* ditunjukkan pada Gambar 6.

	full_text	tokens
0	Dengan dukungan dana 4 miliar USD dari QIA Dan...	[dukungan, dana, miliar, usd, qia, danantara, ...
1	Danantara akan bangun Mega Proyek hasil efisie...	[danantara, bangun, mega, proyek, hasil, efisie...
2	BPI Danantara tengah menjajaki peluang investa...	[bpi, danantara, menjajaki, peluang, investasi...
3	@ARSIPAJA Buat danantara bisa cepat? Itu ngump...	[danantara, cepat, ngumpulin, dana, ratusan, r...
4	Danantara bekerja sama dengan Qatar Investment...	[danantara, qatar, investment, authority, qia,...
5	@Not_Ur_Bro @chaatimeeee @bembengkedua @jinjam...	[sesakti, sih, parameter, lu, pemerintah, pusa...
6	@tempodotco Urusan sekolah layak 30 thn? Tp yg...	[urusan, sekolah, layak, knapa, prioritas, mba...

Gambar 6. Hasil *Stopword* dan *Stemming*

3.2.4 Pelabelan Data

Pada penelitian ini, pelabelan dilakukan menggunakan pendekatan berbasis lexicon Bahasa Indonesia, yaitu dengan mencocokkan kata-kata dalam *tweet* dengan daftar yang memiliki nilai sentimen tertentu. *Lexicon* berisi

kumpulan kata-kata yang telah diberi skor sentimen. Hasil untuk pelabelan berbasis *Lexicon* dapat dilihat pada Gambar 7.

	tokens	polarity_score	polarity
0	[dukungan, 'dana, 'miliar, 'usd, 'qia, '...	11	1
1	[danantara, 'bangun, 'mega, 'proyek, 'has...	6	1
2	[bpi, 'danantara, 'menjajaki, 'peluang, '...	4	1
3	[danantara, 'cepat, 'ngumpulin, 'dana, 'r...	-2	-1
4	[danantara, 'qatar, 'investment, 'authorit...	15	1
5	[sesakti, 'sih, 'parameter, 'lu, 'pemerin...	-6	-1
6	[urusan, 'sekolah, 'layak, 'knapa, 'prior...	9	1

Gambar 7. Hasil Pelabelan menggunakan Lexicon

3.3. TF-IDF

Pada penelitian ini, pelabelan dilakukan menggunakan pendekatan berbasis lexicon Bahasa Indonesia, yaitu dengan mencocokkan kata-kata dalam tweet dengan daftar yang memiliki nilai sentimen tertentu. Lexicon berisi kumpulan kata-kata yang telah diberi skor sentimen pada gambar 8.

	tokens	polarity_score	polarity
0	[dukungan, 'dana, 'miliar, 'usd, 'qia, '...	11	1
1	[danantara, 'bangun, 'mega, 'proyek, 'has...	6	1
2	[bpi, 'danantara, 'menjajaki, 'peluang, '...	4	1
3	[danantara, 'cepat, 'ngumpulin, 'dana, 'r...	-2	-1
4	[danantara, 'qatar, 'investment, 'authorit...	15	1
5	[sesakti, 'sih, 'parameter, 'lu, 'pemerin...	-6	-1
6	[urusan, 'sekolah, 'layak, 'knapa, 'prior...	9	1

Gambar 8. Hasil Pelabelan dengan Lexicon

Setelah data selesai diberikan label, langkah selanjutnya dalam penelitian ini adalah menerapkan pembobotan kata menggunakan metode TF-IDF (*Term Frequency – Inverse Document Frequency*). Metode ini digunakan untuk mengukur seberapa penting sebuah kata dalam satu dokumen menggunakan *TfidfVectorizer* dari *Scikit-Learn*. Tujuan dari prosesi ini adalah untuk menekankan kata-kata yang sering muncul dalam suatu tweet namun jarang muncul didalam tweet lainnya. Hasil untuk pembobotan menggunakan TF-IDF dapat dilihat pada Gambar 9.

```

Kata: proyek, Kemunculan: 110.62036807511168
Kata: investasi, Kemunculan: 80.73721759646487
Kata: kelola, Kemunculan: 77.90173626821904
Kata: prabowo, Kemunculan: 77.29066446305822
Kata: daya, Kemunculan: 67.26135545638195
Kata: nusantara, Kemunculan: 65.72863457780625
Kata: anagata, Kemunculan: 64.33684037653467
Kata: anagata nusantara, Kemunculan: 63.905870028445854
Kata: daya anagata, Kemunculan: 63.88020235489845
Kata: presiden, Kemunculan: 63.84309466096241
    
```

Gambar 9. Hasil TF-IDF

3.4. SMOTE (*Synthetic Minority Over-Sampling Technique*)

Setelah data dibagi menjadi data latih dan data uji, dilakukan analisis terhadap distribusi label pada data latih. Hasil analisis menunjukkan adanya ketidakseimbangan jumlah data antar kelas sentimen positif 1.491 dan Negatif 1.256 sehingga di seimbangkan jumlahnya positif dan negatif 1.256.

3.5. Klasifikasi dengan Metode SVM (*Support Vector Machine*)

Setelah data dibagi dan diseimbangkan, proses klasifikasi dilakukan dengan menggunakan *Support Vector Machine* (SVM). Dalam penelitian ini, model SVM digunakan untuk mengklasifikasikan sentimen dari data ulasan terkait proyek Danantara yang dikumpulkan dari media sosial X. Berdasarkan hasil uji performa dengan menggunakan model SVM yang dioptimalkan menggunakan kernel RBF diperoleh nilai *Accuracy* sebesar 87.25%, *Precision* sebesar 87.49%, *Recall* sebesar 87.25%, dan *f1 score* sebesar 87.23%. Hal ini menunjukkan bahwa algoritma SVM mampu mengenali pola sentimen dalam data secara cukup efektif.

3.6. Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa model klasifikasi *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen dari data tweet. Proses evaluasi dilakukan menggunakan *confusion matrix*, serta metrik evaluasi berupa *accuracy*, *precision*, *recall* dan *f1-score*.



Gambar 10. *Confusion Matrix*

Berdasarkan *confusion matrix* pada gambar 10, dapat disimpulkan bahwa:

1. Model berhasil mengklasifikasikan 168 dari 189 data negatif dengan benar.
2. Model cukup baik dalam mengklasifikasikan data positif, yaitu 192 dari 224 data berhasil diprediksi dengan benar.
3. Terdapat 21 data negatif yang salah diklasifikasikan sebagai positif, dan 32 data positif yang diklasifikasikan sebagai negatif.

Sedangkan pada matrik evaluasi mendapatkan hasil:

1. Akurasi model sebesar 87%.
2. Recall pada kedua kelas relatif seimbang pada 86%.
3. Precision model sebesar 90%, menunjukkan bahwa prediksi positif yang dihasilkan model sebagian besar benar.

Model SVM menunjukkan performa yang kuat dan seimbang dalam mengklasifikasikan sentimen positif dan negatif. Hasil ini mengindikasikan bahwa model mampu mengenali pola sentimen secara efektif pada data yang telah diproses.

4. KESIMPULAN

Berdasarkan hasil penelitian mengenai analisis sentimen masyarakat Indonesia terhadap peluncuran proyek Danantara melalui media sosial X dengan menggunakan algoritma *Support Vector Machine* (SVM) bahwa Sentimen positif terlihat lebih dominan secara kuantitatif, namun tidak dapat diabaikan bahwa terdapat pula sentimen negatif yang cukup signifikan. Sentimen positif umumnya berkaitan dengan harapan terhadap kemajuan investasi nasional dan pengelolaan strategis proyek, sementara sentiment negatif banyak menyoroti isu korupsi, ketidaktransparan, dan kekhawatiran terhadap potensi kegagalan proyek. Sebagaimana tercermin dalam wordcloud kata-kata seperti korupsi, gagal, dan rugi. Kategori sentiment yang paling dominan adalah sentiment positif. Dari hasil *confusion matrix*, terdapat 192 data yang terklasifikasi sebagai sentimen positif secara benar (*true positive*), lebih banyak dibandingkan sentimen *negative* (168 *true negative*). Meskipun demikian, temuan ini tidak sepenuhnya mencerminkan penerimaan yang mutlak terhadap proyek, karena sebagian sentiment positif teridentifikasi berasal dari akun-akun yang berperilaku seperti buzzer. Di sisi lain, sentimen negative mengandung kritik terkait aspek pengelolaan, potensi korupsi, dan efisiensi dana publik. Efektivitas algoritma *Support Vector Machine* (SVM) dalam klasifikasi sentimen publik terhadap proyek Danantara menunjukkan hasil yang sangat baik. Berdasarkan evaluasi terhadap 3152 data yang diambil pada periode 24 Februari hingga 4 April 2025, model SVM mampu mencapai akurasi sebesar 87%, precision sebesar 86%, recall sebesar 90%, dan F1-score sebesar 88%. Hasil analisis sentimen ini memberikan gambaran awal yang cukup representatif terhadap penerimaan publik terhadap proyek Danantara. Secara umum, masyarakat menunjukkan ekspektasi tinggi terhadap proyek ini, namun juga menyimpan kekhawatiran besar terkait pelaksanaan dan transparansi. Oleh karena itu, persepsi publik terhadap proyek ini masih berada pada tahap yang fluktuatif dan perlu ditindaklanjuti dengan komunikasi publik dan pengelolaan risiko yang baik oleh pihak pemerintah.

REFERENSI

- [1] "Daya Anagata Nusantara 'Untuk Kemakmuran Indonesia,'" Danantara Indonesia. [Online]. Available: <https://www.danantaraindonesia.com/>

- [2] D. R. Rohmania and R. Abidin, "Analisis Sentimen Masyarakat Indonesia Terhadap Kebijakan Program Tapera Menggunakan Brand24," vol. 1, no. 2, pp. 120–131, 2024.
- [3] M. Alfarizi, M. Rizqy, R. I. Ghufroni, D. Fathurahman, R. D. Saputra, and F. Kurniawan, "Analisis Sentimen Persepsi Publik Terhadap Kasus Bullying Siswa Cilacap Menggunakan Pendekatan Machine Learning," J. Inf. Technol. Ampera, vol. 4, no. 3, pp. 265–276, 2023.
- [4] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," Jambura J. Electr. Electron. Eng., vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [5] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," J. Teknol. Inf. dan Ilmu Komput., vol. 8, no. 1, p. 147, 2021, doi:10.25126/jtiik.0813944.
- [6] C. F. Alifa and D. Alita, "Analisis Opini Publik Tentang Boikot Produk Pro-Israel di Twitter Berbahasa Indonesia Menggunakan Metode SVM," J. Inform. J. Pengemb. IT, vol. 9, no. 2, pp. 112–120, 2024, doi: 10.30591/jpit.v9i2.6559.
- [7] Shahmirul Hafizullah Imanuddin, Kusworo Adi, and Rahmat Gernowo, "Sentiment Analysis on Satusehat Application Using Support Vector Machine Method," J. Electron. Electromed. Eng. Med. Informatics, vol. 5, no. 3, pp. 143–149, 2023, doi: 10.35882/jeemi.v5i3.304.
- [8] W. A. Prabowo and F. Azizah, "Sentiment Analysis for Detecting Cyberbullying Using TF-IDF and SVM," J. RESTI (Rekayasa Sist. Dan Teknol. Informasi), vol. 4, no. 6, pp. 11–12, 2020, doi:10.29207/resti.v4i6.2753.
- [9] I. W. D. Wiguna, D. V. Waas, I. K. A. G. Wiguna, and M. L. Radhitya, "Comparison of SVM & Naïve Bayes Methods in Sentiment Analysis of Electric Vehicle Subsidy Policy Based on X Data," J. Eng. Sci. Res., vol. 6, no. 1, pp. 23–30, 2024, doi: 10.23960/jesr.v6i1.158.
- [10] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," J. RESTI (RekayasaSist. dan Teknol. Informasi), vol. 5, no. 6, pp. 1120–1126, 2021, doi:10.29207/resti.v5i6.3588.
- [11] Husni, A. Muntasa, and M. D. Hartanto, "Classification of Public Opinion Online Learning Policies using Various Support Vector Machine's Kernel," Classif. Public Opin. Online Learn. Policies using Var. SupportVector Mach. Kernel, vol. XIII, no. 3, pp. 254–260, 2023.
- [12] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," J. Inform. J. Pengemb. IT, vol. 6, no. 3, pp. 150–155, 2021.
- [13] D. Retnoningrum, D. A. Putri, I. Budi, A. B. Santoso, and P. K. Putra, "Stance Analysis of Policies Related to Emission Test Obligations using Twitter Social Media Data," J. Nas. Pendidik. Tek. Inform., vol. 12, no. 3, pp. 472–481, 2024, doi: 10.23887/janapati.v12i3.69004.
- [14] G. Machado, "Web Scraping: Data search and retrieval methodology in information science," Acad. Educ. Navig. Path Knowl., 2023, doi:10.56238/sevned2023.008-002.
- [15] R. Bhaskaran et al., "Intelligent Machine Learning with Metaheuristics Based Sentiment Analysis and Classification," Comput. Syst. Sci. Eng., vol. 44, no. 1, pp. 235–247, 2022, doi: 10.32604/csse.2023.024399.
- [16] K. P. Gunasekaran, "Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review," pp. 1–6, 2023.
- [17] F. Marisa, "Data Mining Data mining," Min. Massive Datasets, vol. 2, no. January 2013, pp. 5–20, 2005, [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- [18] N. Rosário-Ferreira et al., "The Treasury Chest of Text Mining: Piling Available Resources for Powerful Biomedical Text Mining," BioChem, vol. 1, no. 2, pp. 60–80, 2021, doi: 10.3390/biochem1020007.
- [19] Ajay Jadhav, Pranjal Jagtap, Suraj Gurav, Shivani Jadhav, Nikita Jadhav, and Afsha Akkalkot, "A Survey on Text Mining - Techniques, Application," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 3307, pp. 338–343, 2023, doi: 10.32628/cseit3307391.
- [20] D. Varma, A. Nehansh, and P. Swathy, "Data Preprocessing Toolkit : An Approach to Automate Data Preprocessing," Interantional J. Sci. Res. Eng. Manag., vol. 07, no. 03, pp. 1–5, 2023, doi: 10.55041/ijrsrem18270.
- [21] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," J. Media Inform. Budidarma, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [22] N. B. Aji, M. Abdi, and A. Rakhmadi, "Tf*Idf and Random Walk For Term Candidate Selection On Automatic Subject Indexing," Jaict, vol. 6, no. 1, p. 38, 2021, doi: 10.32497/jaict.v6i1.2436.
- [23] A. Addiga and S. Bagui, "Sentiment Analysis on Twitter Data Using Term Frequency-Inverse Document Frequency," J. Comput. Commun., vol. 10, no. 08, pp. 117–128, 2022, doi: 10.4236/jcc.2022.108008.
- [24] V. Piccialli and M. Sciandrone, "Nonlinear optimization and support vector machines," Ann. Oper. Res., vol. 314, no. 1, pp. 15–47, 2022, doi: 10.1007/s10479-022-04655-x.
- [25] M. Shamsi and S. Beheshti, "Separability and Scatteredness (S&S) Ratio-Based Efficient SVM Regularization Parameter, Kernel, and Kernel Parameter Selection," pp. 1–21, 2023, [Online]. Available: <http://arxiv.org/abs/2305.10219>
- [26] <http://arxiv.org/abs/2305.10219>
- [27] M. J. Nelson and A. K. Hoover, "Notes on Using Google Laboratory in AI Education," Annu. Conf. Innov. Technol. Comput. Sci. Educ. ITiCSE, pp. 533–534, 2020, doi: 10.1145/3341525.3393997.
- [28] D. Sanmartín Carrión, V. Prohaska, and O. Diez, "Exploration of TPUs for AI Applications," Lect. Notes Networks Syst., vol. 956 LNNS, no. Figure 1, pp. 559–559, 2024, doi: 10.1007/978-3-031-56950-0_47.
- [29] Y. Zahidi, Y. El Younoussi, and Y. Al-Amrani, "Different valuable tools for Arabic sentiment analysis: a comparative evaluation," Int. J. Electr. Comput. Eng., vol. 11, no. 1, pp. 753–762, 2021, doi: 10.11591/ijece.v11i1.pp753-762.
- [30] G. Umadevi, "SENTIMENT ANALYSIS WITH PYTHON : A HANDS-ON APPROACH," Magestic Technol. Solut. Ltd Chennai. Tamil Nadu, India., 2023.
- [31] J. M. Nápoles-Duarte, A. Biswas, M. I. Parker, J. P. Palomares-Baez, M. A. Chávez-Rojo, and L. M. Rodríguez-Valdez, "Stmol: A component for building interactive molecular visualizations within streamlit web-applications," Front. Mol. Biosci., vol. 9, no. September, pp. 1–10, 2022, doi: 10.3389/fmolb.2022.99