

Analisis Sentimen Publik terhadap Kebijakan Danantara Menggunakan Algoritma Support Vector Machine (SVM) pada Komentar di Media Sosial X

Kiki Nurnakiyah¹, Sahara Syarifatul Choeriyah^{1,*}, Natya Taniarza¹, Muhammad Rizky Munggaran¹, Fanisya Dwiyan¹

¹Program Studi Informatika, Sekolah Tinggi Teknologi Cipasung

*Corresponding author's email: sahara@sttcipasung.ac.id

Diterima: 21-08-2026

Direvisi: 28-08-2026

Diterima setelah revisi: 28-08-2026

Abstract. Social media has become a primary platform for the public to express opinions on various issues, including government policies such as Danantara. This policy has sparked both support and criticism from academics as well as the general public. This study aims to analyze public sentiment and evaluate the performance of the Support Vector Machine (SVM) algorithm regarding the Danantara policy based on comments collected from social media X between February and May 2025. The results show that, using a dataset of 1,626 comments and the SVM algorithm, 71.47% of the comments were classified as positive sentiment, while 28.52% were classified as negative sentiment. Before the application of the Synthetic Minority Over-sampling Technique (SMOTE), the SVM model achieved an accuracy of 78%, with the positive class performing better (precision 0.79, recall 0.94, f1-score 0.86) compared to the negative class (precision 0.72, recall 0.35, f1-score 0.47). After applying SMOTE, accuracy improved to 80%, with the negative class showing significant improvement (precision 0.68, recall 0.58, f1-score 0.62), while the positive class remained stable (precision 0.84, recall 0.89, f1-score 0.86). These findings indicate that public sentiment toward the Danantara policy is predominantly positive, and the application of SMOTE helps the model recognize the minority class without reducing the performance of the majority class, thereby achieving a more balanced classification.

Keywords: Support Vector Machine (SVM), Lexicon-based, Synthetic Minority Over-sampling Technique (SMOTE), Danantara, Social Media X.

Abstrak. Media sosial telah menjadi sarana utama bagi masyarakat untuk menyampaikan opini mengenai berbagai isu, termasuk kebijakan pemerintah salah satunya Danantara. Kebijakan tersebut menuai pro kontra dari kalangan akademisi maupun masyarakat. Penelitian ini bertujuan untuk menganalisis sentimen publik dan mengevaluasi performa algoritma *Support Vector Machine* (SVM) terkait kebijakan Danantara pada komentar di media sosial X yang diambil dari bulan Februari hingga Mei 2025. Hasil penelitian menunjukkan bahwa sentimen publik terhadap kebijakan Danantara dengan menggunakan dataset sebanyak 1.626 komentar menggunakan algoritma *Support Vector Machine* (SVM) diperoleh sebesar 71,47% komentar tergolong dalam sentimen positif dan 28,52% komentar tergolong dalam sentimen negatif. Sebelum penerapan *Synthetic Minority Over-sampling Technique* (SMOTE), model SVM memiliki akurasi 78%, dengan kelas positif menunjukkan kinerja lebih baik yaitu memperoleh nilai *precision* 0,79, *recall* 0,94, *f1-score* 0,86 dibanding kelas negatif dengan nilai *precision* 0,72, *recall* 0,35, *f1-score* 0,47. Setelah penerapan SMOTE, akurasi meningkat menjadi 80%, kelas negatif mengalami peningkatan signifikan dengan nilai *precision* 0,68, *recall* 0,58, *f1-score* 0,62, sementara kelas positif tetap stabil dengan nilai *precision* 0,84, *recall* 0,89, *f1-score* 0,86. Hasil ini menunjukkan bahwa sentimen publik terhadap kebijakan Danantara didominasi opini positif dan penerapan SMOTE membantu model mengenali kelas minoritas tanpa menurunkan performa kelas mayoritas, sehingga klasifikasi menjadi lebih seimbang.

Kata kunci: *Support Vector Machine* (SVM), *Lexicon-based*, *Synthetic Minority Over-sampling Technique* (SMOTE), Danantara, Media Sosial X.

I. PENDAHULUAN

Pertumbuhan ekonomi Indonesia pada triwulan III 2024 tercatat 4,95% *year-on-year*, namun masih dihadapkan pada tantangan global seperti fluktuasi harga komoditas, kenaikan suku bunga, gangguan rantai pasokan, serta isu ketahanan pangan dan energi [1]. Untuk memperkuat ketahanan ekonomi dan mendukung visi Indonesia Emas 2045, pemerintah meluncurkan Danantara, sebuah lembaga investasi yang mengelola aset strategis negara guna meningkatkan efisiensi, menarik investasi, dan memperkuat daya saing nasional

[1], [2]. Lembaga ini melibatkan tujuh BUMN besar dengan aset diperkirakan mencapai 900 miliar dolar AS, menjadikannya salah satu dana kekayaan negara terbesar di dunia [4], [5].

Namun, kehadiran lembaga ini tidak lepas dari sorotan publik. Di media sosial X yang banyak digunakan masyarakat Indonesia untuk mengekspresikan opini, baik berupa pujian, kritik, ujaran kebencian, maupun hoaks [6], kebijakan ini memicu beragam respon misalnya, Clara Puspita beropini “Danantara Indonesia percepatan perkembangan Indonesia lebih baik.” Sebaliknya, Jaki mengkritik “jika pajak adalah cara biadab untuk mempertahankan peradaban, maka Danantara adalah perampokan brutal yang dilegalkan demi logistik mempertahankan kekuasaan”. Selain itu, kebijakan Danantara juga menuai sorotan dari kalangan akademisi dan pakar yaitu Charles Simabura, pakar Hukum Tata Negara dari Universitas Andalas, menilai bahwa partisipasi mantan presiden dalam Dewan Penasihat berisiko mengalihkan fokus lembaga dari tujuan ekonomi ke arah kepentingan politik, yang seharusnya dihindari [7]. Kondisi ini menimbulkan pro dan kontra serta keraguan terhadap transparansi dan efektivitas program.

Untuk memahami persepsi publik secara sistematis, digunakan analisis sentimen, yaitu proses yang bertujuan untuk mengidentifikasi, mengevaluasi, dan menentukan apakah suatu teks mengandung sentimen positif, negatif, atau netral [8], [9]. Salah satu algoritma yang digunakan dalam klasifikasi teks adalah *Support Vector Machine* (SVM), yaitu mencari *hyperplane* optimal yang mampu memisahkan dua kelompok data dengan margin maksimum [10].

Berbagai penelitian telah membuktikan keunggulan SVM. Referensi [3] memperoleh akurasi 96,68% pada analisis sentimen pemindahan ibu kota. Referensi [10] memperoleh akurasi 0,97 pada ulasan drama Korea, lebih tinggi dari KNN (0,95). Referensi [11] juga menunjukkan SVM unggul dalam analisis kebijakan makan gratis dengan akurasi tertinggi dalam analisis kebijakan makan gratis di Indonesia, mencapai 98% sebelum dan meningkat menjadi 99% setelah penerapan SMOTE. *Decision Tree* mencatat akurasi 98%, sementara *Naive Bayes* memiliki akurasi terendah, yaitu 92% sebelum dan 91% setelah SMOTE.

Perbedaan utama penelitian ini terletak pada topik yang dibahas, yaitu kebijakan Danantara yang belum dianalisis dalam penelitian sentimen. Penelitian ini bertujuan untuk menganalisis sentimen publik dan mengevaluasi performa algoritma *Support Vector Machine* (SVM) terkait kebijakan Danantara pada komentar di media sosial X. Hasil yang diperoleh diharapkan dapat menggambarkan respons publik secara nyata dan menunjukkan seberapa baik algoritma SVM bekerja dalam analisis sentimen kebijakan Danantara.

II. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen merupakan salah satu metode dalam bidang *Natural Language Processing* (NLP) yang berfungsi untuk mendeteksi, mengevaluasi, serta mengklasifikasikan emosi atau opini yang terkandung dalam suatu teks [15]. Data yang dianalisis dalam proses ini biasanya berasal dari berbagai sumber, seperti ulasan, komentar, cuitan, atau postingan di media sosial [10].

Tujuan utama analisis sentimen adalah untuk menentukan apakah suatu teks mengandung sentimen positif, negatif, atau netral [8]. Proses ini dilakukan dengan memanfaatkan teknik analisis teks dan algoritma komputasi yang dirancang untuk mengenali pola serta makna dalam kalimat, sehingga memungkinkan pengelompokan opini masyarakat secara sistematis [9].

2.2 Preprocessing Data

Preprocessing data merupakan serangkaian proses yang bertujuan untuk mempersiapkan dan membersihkan data mentah sebelum digunakan dalam model atau algoritma analisis. Ada beberapa tahapan yang dilakukan dalam *preprocessing* [13].

1. *Cleaning* adalah proses membersihkan data teks seperti simbol, link URL, *hashtag*, *emoticon* dan *mention*(@) yang tidak diperlukan. Bertujuan agar teks lebih bersih dan siap untuk diproses.
2. *Case Folding* merupakan proses mengubah semua huruf dalam teks menjadi huruf kecil. Bertujuan untuk menyamakan bentuk kata sehingga perbedaan huruf kapital tidak mempengaruhi analisis.
3. *Normalization* merupakan proses untuk memperbaiki kata tidak baku menjadi kata baku.
4. *Tokenizing* merupakan tahapan untuk memecah teks atau kalimat menjadi unit-unit yang lebih kecil.
5. *Stopword Removal* adalah proses menghapus kata-kata umum yang sering muncul tetapi tidak memiliki makna signifikan dalam analisis, seperti dan, di, ke, yang.
6. *Stemming* adalah proses mengubah kata-kata pada data yang memiliki kata imbuhan menjadi kata dasar.

2.3 Support Vector Machine (SVM)

Support Vector Machine pertama kali diperkenalkan oleh Vapnik, Boser dan Guyon serta dikembangkan oleh Boser pada tahun 1992 [14]. Algoritma ini adalah model *Machine Learning* termasuk ke dalam kategori *supervised learning* yang biasanya digunakan untuk klasifikasi dan regresi [15]. Konsep klasifikasi dengan

support vector machine yaitu mencari *hyperplane* optimal yang mampu memisahkan dua kelompok data dengan margin maksimum, di mana *hyperplane* berfungsi sebagai batas pemisah antara dua kelas data, *margin* merupakan jarak antara kedua kelompok terhadap *hyperplane*, dan *support vector* adalah titik data terdekat dengan *hyperplane* yang menentukan posisi optimalnya [10]. Referensi [15] *Support Vector Machine* dibagi menjadi dua jenis yaitu:

1. *Support Vector Machine Linear* digunakan untuk data yang bisa dipisahkan dengan garis lurus. Jika suatu *dataset* dapat diklasifikasikan menjadi dua kelompok hanya dengan satu garis pemisah, maka data tersebut disebut *linear separable*, dan model yang digunakan disebut *Linear SVM classifier*.
2. *Support Vector Machine Nonlinear* digunakan untuk data yang tidak bisa dipisahkan dengan lurus. Untuk mengatasi hal ini, trik *kernel* digunakan untuk mengubah data ke dimensi yang lebih tinggi dengan bantuan fungsi matematika, sehingga kelas-kelas dalam data dapat dipisahkan secara *linier*.

Berikut merupakan fungsi *kernel* yang populer dan sering digunakan antara lain sebagai berikut [16].

1. *Linear kernel* merupakan fungsi *kernel* paling sederhana yang digunakan ketika data dapat dipisahkan secara *linier*. *Kernel* ini efektif untuk *dataset* dengan jumlah fitur yang besar karena tidak membutuhkan transformasi ke dimensi lebih tinggi untuk meningkatkan performa, seperti pada klasifikasi teks.
2. *Polynomial kernel* merupakan fungsi *kernel* yang digunakan ketika data tidak terpisah secara *linear*. *Polynomial kernel* sangat cocok untuk permasalahan dimana semua *training dataset* dinormalisasi.
3. *RBF kernel* merupakan fungsi *kernel* yang biasa digunakan dalam analisis ketika data tidak terpisah secara *linear*. *RBF kernel* memiliki dua parameter yaitu *Gamma* dan *Cost*.

2.4 Term Frequency - Inverse Document Frequency (TF-IDF)

TF-IDF adalah teknik yang banyak digunakan dalam pemrosesan bahasa alami (NLP) dan *Machine Learning* untuk menganalisis data teks. Teknik ini digunakan untuk mengubah kumpulan dokumen teks menjadi angka atau vektor numerik, sehingga bisa diproses oleh algoritma pembelajaran mesin [13]. Cara kerja TF-IDF memiliki dua komponen utama yaitu.

1. *Term Frequency* (TF) – Menghitung seberapa sering suatu kata muncul dalam dokumen. Semakin sering kata muncul dalam dokumen, semakin besar nilainya.

$$TF = \frac{\text{JUMLAH KEMUNCULAN KATA DALAM DOKUMEN}}{\text{TOTAL KATA DALAM DOKUMEN}} \quad (2.1)$$

2. *Inverse Document Frequency* (IDF) – Menghitung seberapa unik kata tersebut dalam keseluruhan dokumen.

$$IDF = \log \left(\frac{\text{TOTAL JUMLAH DOKUMEN}}{\text{JUMLAH DOKUMEN YANG MENGANDUNG KATA TERSEBUT}} \right) \quad (2.2)$$

Kata yang sering muncul di banyak dokumen memiliki IDF rendah, sedangkan kata yang jarang muncul memiliki IDF tinggi, menandakan kata itu lebih penting.

TF-IDF dihitung dengan mengalikan kedua nilai tersebut:

$$TF - IDF = TF \times IDF \quad (2.3)$$

2.5 Confusion Matrix

Confusion matrix merupakan tabel yang menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang diklasifikasikan dengan salah. Menunjukkan informasi tentang perbandingan hasil klasifikasi yang dilakukan sistem (prediksi) dengan hasil klasifikasi yang sebenarnya. *Confusion matrix* terdapat empat hasil nilai klasifikasi yaitu *True Positive* (TP) adalah ketika model memprediksi positif dan hasilnya memang benar. *False Positive* (FP) terjadi saat model memprediksi positif, tetapi sebenarnya salah. *False Negative* (FN) adalah saat model memprediksi negatif padahal seharusnya positif. *True Negative* (TN) terjadi ketika model memprediksi negatif dan hasilnya memang benar [8]. Dengan menggunakan *Confusion Matrix*, dapat dihitung sejumlah indikator evaluasi seperti *accuracy*, *precision*, *recall*, dan *f1-score* untuk menilai kinerja model dalam mengklasifikasi data [14].

Accuracy adalah menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (2.4)$$

Precision menunjukkan seberapa akurat model dalam memprediksi data yang sesuai dengan yang diharapkan.

$$Precision = \frac{TP}{TP+FP} \quad (2.5)$$

Recall mengukur kemampuan model dalam mengenali dan menemukan kembali informasi yang relevan.

$$Recall = \frac{TP}{TP+FN} \quad (2.6)$$

F1-score merupakan rata-rata dari presisi dan *recall*, yang berfungsi untuk menciptakan keseimbangan antara keduanya.

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (2.7)$$

2.6 Crawling Data

Crawling data adalah teknik otomatis untuk mengumpulkan informasi dari internet dengan bantuan program yang dinamakan *web crawler*. Data yang telah dikumpulkan kemudian disimpan dan dapat digunakan untuk berbagai keperluan, seperti analisis. Cara kerja *crawling* data yaitu *web crawler* akan menjelajahi setiap URL yang diberikan, dan mengunjungi URL tersebut secara satu per satu. Setiap halaman URL akan diidentifikasi apakah URL tersebut sudah masuk ke *database*. Jika ada URL yang belum masuk maka akan ditambahkan kedalam *database*. Siklus mengambil dan menambahkan list URL ini akan terus menerus dan berlanjut [17].

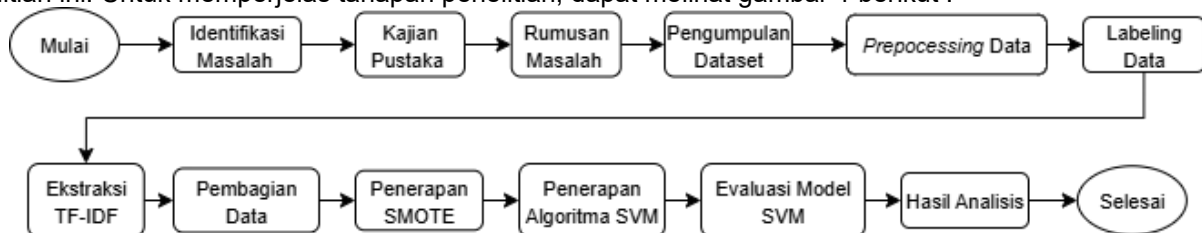
2.7 Synthetic Minority Over-sampling Technique (SMOTE)

Synthetic Minority Over-sampling Technique (SMOTE) adalah metode untuk mengatasi ketidakseimbangan kelas dalam dataset ketika jumlah data minoritas jauh lebih sedikit dibandingkan mayoritas. Teknik ini bekerja dengan menambahkan data sintesis di kelas minoritas berdasarkan kedekatan dengan tetangga terdekatnya. Pendekatan tersebut membantu menyeimbangkan distribusi data sehingga model tidak bias terhadap kelas mayoritas dan mampu menghasilkan prediksi yang lebih akurat [8].

III. METODE PENELITIAN

3.1 Tahapan Penelitian

Bagian ini menjelaskan mengenai tahapan yang akan dilakukan dalam menyusun dan menyelesaikan penelitian ini. Untuk memperjelas tahapan penelitian, dapat melihat gambar 1 berikut :



Gambar 1 Tahapan Penelitian

1. Identifikasi masalah merupakan tahap awal penelitian untuk memahami fenomena, menilai ketersediaan data, serta menentukan arah dan fokus penelitian.
2. Kajian pustaka dilakukan untuk memahami teori dan penelitian terdahulu yang relevan sebagai dasar dalam menentukan arah dan metode penelitian.
3. Perumusan Masalah, yaitu menetapkan fokus penelitian, tujuan, algoritma yang digunakan, serta tools pendukung analisis sentimen.
4. Pengumpulan *Dataset*, data diambil dari komentar masyarakat di media sosial X terkait kebijakan Danantara melalui crawling dengan X API dan Tweet-Harvest berdasarkan kata kunci #dantara.
5. *Preprocessing* Data, meliputi *cleaning*, *case folding*, *normalization*, *tokenizing*, *stopword removal*, dan *stemming* agar data siap diproses.
6. Labeling Data, dilakukan menggunakan metode *lexicon-based* untuk mengklasifikasikan komentar menjadi sentimen positif dan negatif.
7. Ekstraksi Fitur (TF-IDF), yaitu mengubah teks menjadi bentuk numerik agar dapat diproses oleh algoritma Machine Learning.
8. Pembagian Data, *dataset* dibagi menjadi 80% data latih dan 20% data uji untuk pelatihan dan pengujian model.
9. Penerapan SMOTE, digunakan untuk menyeimbangkan jumlah data antar kelas agar model tidak bias.
10. Penerapan Algoritma SVM., SVM digunakan untuk mengklasifikasikan komentar menjadi sentimen positif dan negatif berdasarkan pola data latih.
11. Evaluasi Model, kinerja model diukur menggunakan *Confusion Matrix* dengan indikator *accuracy*, *precision*, *recall*, dan *f1-score*.
12. Analisis Hasil, menganalisis performa model serta menggambarkan persepsi publik terhadap kebijakan Danantara berdasarkan hasil sentimen.

3.3 Analisis Data

Analisis data pada penelitian ini diawali dengan proses pelabelan menggunakan metode *lexicon-based* untuk menentukan kategori sentimen positif dan negatif. Selanjutnya dilakukan *preprocessing* teks, ekstraksi fitur dengan TF-IDF, serta penerapan SMOTE guna menyeimbangkan distribusi kelas. Algoritma Support Vector Machine (SVM) diimplementasikan menggunakan library scikit-learn pada platform Google Colab. Evaluasi kinerja model dilakukan dengan *confusion matrix*, yang digunakan untuk menghitung indikator *accuracy*, *precision*, *recall*, dan *f1-score*. Hasil evaluasi memberikan gambaran menyeluruh mengenai kemampuan model dalam mengklasifikasikan komentar sesuai kategori sentimen.

IV. HASIL DAN PEMBAHASAN

4.1 Hasil

4.1.1 Pengumpulan Dataset

Pengumpulan *dataset* dilakukan dengan mengambil komentar masyarakat di media sosial X terkait kebijakan Danantara. Proses *crawling* menggunakan X API melalui *tool Tweet-Harvest* di Google Collab dengan kata kunci #dantara, data diambil pada periode bulan Februari–Maret 2025 dan terkumpul 1.891 data.

4.1.2 Preprocessing Data

Preprocessing data merupakan serangkaian proses yang bertujuan untuk mempersiapkan dan membersihkan data mentah sebelum digunakan dalam model atau algoritma analisis. Ada beberapa tahapan yang dilakukan dalam preprocessing (Yuspita & Suryono, 2024).

1. *Cleaning* merupakan proses pembersihan data teks dengan menghapus simbol, URL, emotikon, *mention* (@), serta data duplikat menggunakan fungsi *drop_duplicates()* agar teks lebih bersih dan siap diproses. Hasil pembersihan mengurangi jumlah data dari 1.891 menjadi 1.626, seperti ditunjukkan pada Tabel 4.1.

Tabel 4.1 Hasil Proses *Cleaning*

<i>Full_text</i>	<i>Cleaning_text</i>
@probowo Penjelasan simpel DANANTARA https://t.co/6dquWZebQi	Penjelasan simpel DANANTARA
Apa sih Danantara itu? https://t.co/NH8cVuNMRS	Apa sih Danantara itu

2. *Case folding*, yaitu proses mengubah semua huruf dalam teks menjadi huruf kecil. Bertujuan untuk menyamakan bentuk kata sehingga perbedaan huruf kapital tidak mempengaruhi analisis. Hasilnya disimpan pada kolom *case_folding* dan ditunjukkan pada Tabel 4.2.

Tabel 4.2 Hasil Proses *Case Folding*

<i>Cleaning</i>	<i>Case_folding</i>
Penjelasan simpel DANANTARA	penjelasan simpel danantara
Apa sih Danantara itu	apa sih danantara itu

3. *Normalization*, yaitu tahapan untuk mengubah kata-kata tidak baku menjadi bentuk baku sesuai ejaan bahasa Indonesia. Proses ini penting karena data dari media sosial sering menggunakan bahasa informal dan singkatan. Hasilnya disimpan pada kolom *normalize_text* dan ditunjukkan pada Tabel 4.3.

Tabel 4.3 Hasil *Normalization*

<i>Case_folding</i>	<i>Normalize_text</i>
kalau kalian kesel sama danantara tunggu dulu baca yang mimin ringkasin ttg danantara buat kalian awas nambah kesel	kalau kalian kesel sama danantara tunggu dulu baca yang mimin ringkasin tentang danantara buat kalian awas nambah kesel
gak usah bingung lagi danantara terpercaya selalu ada buat bantu capai impian lo danantara bangun negeri	tidak usah bingung lagi danantara terpercaya selalu ada buat bantu capai impian lo danantara bangun negeri

4. *Tokenizing*, yaitu tahapan untuk memecah kalimat menjadi unit-unit yang lebih kecil. Tujuan dari proses ini adalah agar setiap kata dapat dianalisis secara terpisah. Hasilnya disimpan di kolom

tokens dan ditunjukkan pada Tabel 4.4.

Tabel 4.4 Hasil *Tokenizing*

<i>Normalize_text</i>	<i>Tokens</i>
penjelasan simpel danantara	[penjelasan, simpel, danantara]
apa sih danantara itu	[apa, sih, danantara, itu]

5. *Stopword Removal*, yaitu proses menghapus kata-kata umum yang sering muncul tetapi tidak memiliki makna penting dalam analisis, seperti dan, di, ke, yang. Hasilnya disimpan pada kolom *stopword_removal* dan ditunjukkan pada Tabel 4.5.

Tabel 4.5 Hasil *Stopword Removal*

<i>Tokens</i>	<i>Stopword_removal</i>
[tidak, usah, bingung, lagi, danantara,..]	[usah, bingung, danantara,..]
[apa, sih, danantara, itu]	[danantara]

6. *Stemming*, yaitu proses mengubah kata-kata pada data yang memiliki kata imbuhan menjadi kata dasar. Tujuannya agar kata-kata dengan makna yang sama tidak dianggap berbeda dalam analisis. Hasilnya disimpan pada kolom *stemmed* dan ditunjukkan pada Tabel 4.6.

Tabel 4.6 Hasil *Stemming*

<i>Stopword_remove</i>	<i>Stemmed</i>
[bingung, didirikan, peluang, ancaman,..]	[bingung, diri, peluang, ancam,..]
[usah, bingung, danantara, terpercaya, selalu,..]	[usah, bingung, danantara, percaya, selalu,..]

4.1.3 Labeling Data

Labeling data dilakukan menggunakan metode *lexicon-based*, yaitu metode yang memanfaatkan kamus sentimen berisi daftar kata-kata berlabel positif dan negatif dari fajri91. Setiap komentar dicocokkan dengan kata-kata dalam kamus, kemudian diklasifikasikan menjadi dua kelas: skor ≥ 0 untuk sentimen positif dan skor < 0 untuk sentimen negatif. Skor sentimen dihitung menggunakan rumus:

$$\text{Skor total} = \sum_{i=1}^n \text{bobot}(\text{kata } i) \quad (5.1)$$

Tabel 4.7 Contoh Hasil Pelabelan Sentimen

No	Teks Komentar	Score	Sentimen
1.	[jelas, simpel, danantara]	4	Positif
2.	[kalian, kesel, danantara, tunggu, dulu,..]	-2	Negatif

Tabel 4.7 adalah contoh hasil pelabelan sentimen menggunakan metode *lexicon-based*.

Tabel 4.8 Distribusi Sentimen

No	Labeling	Jumlah
1.	Sentimen Positif	1163
2.	Sentimen Negatif	463

Tabel 4.8 menunjukan informasi dari jumlah data berdasarkan hasil pelabelan sentimen. Sebanyak 1.163 data termasuk dalam sentimen positif, sedangkan 463 data tergolong ke dalam sentimen negatif.



Gambar 2 Wordcloud Sentimen Positif



Gambar 3 Wordcloud Sentimen Negatif

Gambar 2 menunjukkan *wordcloud* pada komentar positif seperti “danantara” dan “indonesia”. Sedangkan Gambar 3 menunjukkan kata pada komentar negatif seperti “danantara” dan “korupsi” yang menggambarkan kritik dan kekhawatiran masyarakat.

4.1.4 Ekstraksi TF-IDF

Ekstraksi TF-IDF digunakan untuk mengubah teks menjadi vektor numerik agar dapat diproses oleh algoritma *Machine Learning*. Setiap kata diberi bobot berdasarkan frekuensi kemunculannya dalam dokumen dan pentingnya kata tersebut dibandingkan seluruh dokumen, sehingga kata yang lebih sering muncul dan lebih penting mendapatkan skor tinggi, sedangkan kata yang umum muncul di banyak dokumen memiliki skor rendah.

--- Analisis Tabel TF-IDF ---

Term	TF (avg)	df (Document Frequency)	idf (Inverse Document Frequency)	avg tf-idf
danantara	1.162977	1599	1.016734	0.062722
kelola	0.186962	261	2.826149	0.026385
bumn	0.181427	256	2.845417	0.022847
investasi	0.166052	235	2.930661	0.022518
indonesia	0.160517	230	2.952075	0.021054
negara	0.138376	199	3.096176	0.019209
ekonomi	0.110086	144	3.417759	0.018794
dana	0.092866	130	3.519296	0.014310
perintah	0.081796	125	3.558211	0.012449
indonesiagelap	0.082411	124	3.566179	0.013102

Gambar 5.16 Analisis Tabel TF-IDF

4.1.5 Pembagian Data

Pembagian ini bertujuan untuk mempersiapkan data agar model dapat dilatih dengan sebagian data yang telah diproses. Data pelatihan yang digunakan yaitu 80% dan 20% untuk data uji. Data pelatihan digunakan untuk melatih model agar dapat mengenali pola dan hubungan dalam data, sementara data pengujian digunakan untuk menguji kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya, sehingga dapat dievaluasi kinerjanya.

4.1.6 Penerapan SMOTE

Pada tahapan ini, digunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan jumlah data antar kelas. SMOTE membuat data sintetis pada kelas minoritas agar jumlahnya seimbang dengan kelas mayoritas.

Tabel 4.9 Perbandingan Hasil Model SVM sebelum dan sesudah SMOTE.

Kondisi Data	Sentimen	Accuracy	Presicion	Recall	F1-Score
Sebelum SMOTE	Positif	78%	0.79	0.94	0.86
	Negatif		0.72	0.35	0.47
Sesudah SMOTE	Positif	80%	0.84	0.89	0.86
	Negatif		0.68	0.58	0.62

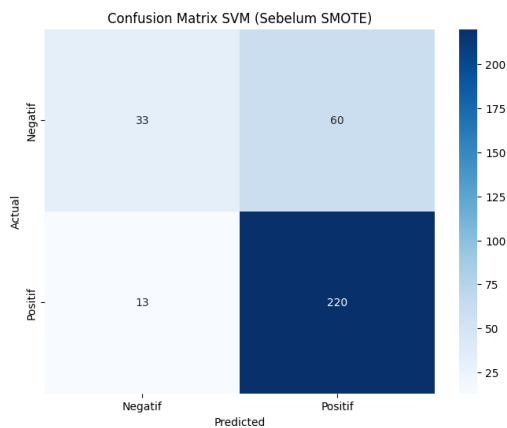
4.1.7 Penerapan Algoritma SVM

Tahap selanjutnya adalah penerapan algoritma *Support Vector Machine* (SVM) sebagai metode klasifikasi pada data yang telah melalui proses pelatihan sebelumnya. Pada tahap ini, digunakan jenis SVM

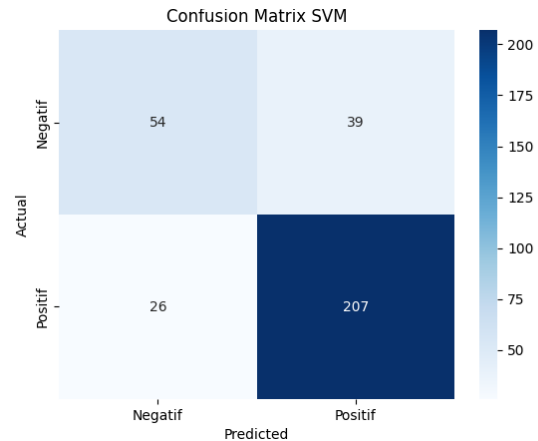
dengan *kernel linier*, yang bertujuan untuk memisahkan data ke dalam dua kelas berdasarkan garis pemisah lurus dalam ruang fitur. Hasil klasifikasi dengan algoritma SVM menunjukkan bahwa sentimen positif (71,47%) sedangkan sentimen negatif (28,52%). Hal ini menggambarkan bahwa tanggapan publik terhadap kebijakan Danantara pada komentar di media sosial X cenderung lebih positif, sehingga dapat disimpulkan kebijakan tersebut memperoleh penerimaan yang lebih besar dibandingkan penolakan dari masyarakat.

4.1.8 Evaluasi Model

Confusion matrix digunakan untuk menggambarkan kinerja model klasifikasi dengan membandingkan antara label prediksi dan label aktual.



Gambar 3 Confusion Matrix sebelum smote



Gambar 4 Confusion Matrix sesudah smote

Gambar 3 dan 4 menunjukkan hasil dari *Confusion Matrix* SVM sebelum dan sesudah SMOTE yaitu:

1. *True Positive* (TP): Sebanyak 220 data sebelum SMOTE dan 207 data sesudah SMOTE data diklasifikasikan sebagai positif oleh model, dan benar-benar termasuk dalam kelas positif
2. *False Positive* (FP): Sebanyak 60 data sebelum SMOTE dan 39 data sesudah SMOTE diklasifikasikan sebagai positif, padahal sebenarnya merupakan data negatif.
3. *True Negative* (TN): Sebanyak 33 data sebelum SMOTE dan 54 data sesudah SMOTE diklasifikasikan sebagai negatif dan memang benar berasal dari kelas.
4. *False Negative* (FN): Sebanyak 13 data sebelum SMOTE dan 26 data sesudah SMOTE diklasifikasikan sebagai negatif, padahal sebenarnya merupakan data positif.

4.2 Pembahasan

Analisis sentimen terhadap kebijakan Danantara dilakukan menggunakan *dataset* yang bersumber dari komentar di media sosial X. Proses analisis diawali dengan tahap *preprocessing* terhadap 1.891 komentar mentah, seperti *cleaning*, *case folding*, dan *tokenizing*, sehingga diperoleh 1.626 data bersih untuk dianalisis. Pelabelan sentimen dilakukan menggunakan metode *lexicon-based*, yang menghasilkan 1.163 komentar positif (71,53%) dan 463 komentar negatif (28,47%). Hasil penerapan model SVM pada *dataset* yang sama menunjukkan distribusi sentimen yaitu 71,47% positif dan 28,52% negatif. Dari analisis ini, kata 'percaya' muncul sebagai contoh pada sentimen positif, sedangkan kata 'korupsi' menjadi contoh pada sentimen negatif. Karena adanya ketidakseimbangan jumlah data antar kelas, dilakukan penyeimbangan menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) sebelum proses pelatihan model.

Sebelum penerapan SMOTE, model SVM memperoleh akurasi sebesar 78%, dengan kelas positif menunjukkan *precision* 0,79, *recall* 0,94, dan *f1-score* 0,86, sedangkan kelas negatif *precision* 0,72, *recall* 0,35, dan *f1-score* 0,47. Rendahnya performa pada kelas negatif disebabkan oleh ketidakseimbangan jumlah data yang membuat model lebih cenderung memprediksi kelas mayoritas. Setelah SMOTE diterapkan, akurasi meningkat menjadi 80%, dengan kelas negatif menunjukkan peningkatan menjadi *precision* 0,68, *recall* 0,58, dan *f1-score* 0,62, sementara kelas positif tetap stabil *precision* 0,84, *recall* 0,89, *f1-score* 0,86. Analisis *confusion matrix* memperlihatkan bahwa *recall* kelas positif masih lebih tinggi dibanding kelas negatif, namun SMOTE berhasil menyeimbangkan prediksi kedua kelas. Selain itu, keterbatasan pelabelan juga mempengaruhi kinerja model, terutama dalam mengenali sentimen negatif. Secara keseluruhan, penerapan SMOTE meningkatkan kemampuan model dalam mendeteksi kelas minoritas tanpa menurunkan performa kelas mayoritas, sehingga klasifikasi menjadi lebih seimbang.

V. KESIMPULAN

Berdasarkan hasil analisis dan rumusan masalah, diperoleh hasil kesimpulan untuk menjawab rumusan masalah tersebut, yaitu:

1. Berdasarkan hasil analisis sentimen publik terhadap kebijakan Danantara dengan menggunakan *dataset* sebanyak 1.626 komentar dari media sosial X menggunakan algoritma *Support Vector Machine* (SVM) diperoleh sebanyak 71,47% komentar tergolong dalam sentimen positif dan 28,52% komentar tergolong dalam sentimen negatif. Hasil ini menunjukkan bahwa opini publik terhadap kebijakan Danantara secara umum cenderung bersifat positif.
2. Evaluasi model algoritma SVM dilakukan dengan terlebih dahulu menerapkan *Synthetic Minority Over-sampling Technique* (SMOTE). Hal ini dilakukan sebagai upaya menyeimbangkan data, sehingga data yang sebelumnya terdiri dari 1163 komentar positif dan 463 komentar negatif menjadi 1163 komentar positif dan 1163 komentar negatif. Sebelum penerapan SMOTE, model SVM memperoleh akurasi sebesar 78%, dengan kelas positif menunjukkan *precision* 0,79, *recall* 0,94, dan *f1-score* 0,86, sedangkan kelas negatif *precision* 0,72, *recall* 0,35, dan *f1-score* 0,47. Rendahnya performa pada kelas negatif disebabkan oleh ketidakseimbangan jumlah data yang membuat model lebih cenderung memprediksi kelas mayoritas. Setelah SMOTE diterapkan, akurasi meningkat menjadi 80%, dengan kelas negatif menunjukkan peningkatan menjadi *precision* 0,68, *recall* 0,58, dan *f1-score* 0,62, sementara kelas positif tetap stabil *precision* 0,84, *recall* 0,89, *f1-score* 0,86. Analisis *confusion matrix* memperlihatkan bahwa *recall* kelas positif masih lebih tinggi dibanding kelas negatif, namun SMOTE berhasil menyeimbangkan prediksi kedua kelas.

DAFTAR PUSTAKA

- [1] A. Hartanto, "Memasuki Tahun 2025, Perekonomian Nasional Diproyeksikan Tetap Terjaga," Kementerian koordinator bidang perekonomian republik indonesia. [Online]. Available: <https://www.ekon.go.id/publikasi/detail/6133/memasuki-tahun-2025-perekonomian-nasional-diproyeksikan-tetap-terjaga>
- [2] Detik.Com, "Resmi! Prabowo Luncurkan Danantara Baca artikel detikbali, 'Resmi! Prabowo Luncurkan Danantara' selengkapnya <https://www.detik.com/bali/bisnis/d-7792522/resmi-prabowo-luncurkan-danantara>. Download Apps Detikcom Sekarang <https://apps.detik.com/detik/>," Detik bali. [Online]. Available: <https://www.detik.com/bali/bisnis/d-7792522/resmi-prabowo-luncurkan-danantara>
- [3] Danantara.id, "Daya Anagata Nusantara untuk kemakmuran indonesia," Danantara indonesia. [Online]. Available: <https://www.danantara.id/>
- [4] CNBCIndonesia.com, "Danantara Resmi Diluncurkan, Tahap Awal Bakal Kelola 7 BUMN Ini," CNBCIndonesia.com. [Online]. Available: <https://www.cnbcindonesia.com/market/20250224110253-17-613025/danantara-resmi-diluncurkan-tahap-awal-bakal-kelola-7-bumn-ini>
- [5] Kompas.id, "Kelola Aset hingga Rp 15.000 Triliun, Nasib Danantara di Tangan Nakhoda Baru," Kompas.id. [Online]. Available: https://www.kompas.id/artikel/kelola-aset-hingga-rp-15000-triliun-nasib-danantara-di-tangan-nakhoda-baru?utm_source=chatgpt.com
- [6] R. A. Hasibuan, D. E. Ratnawati, and R. S. Perdana, "Analisis sentimen kebijakan ekspor pasir laut pada sosial media twitter menggunakan algoritma support vector machine," *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi (JUST-SI)*, vol. 5, no. 1, Aug. 2024, doi: <https://pdfs.semanticscholar.org/ae8b/8b98190df65add51190078e9699d9450f69>.
- [7] R. Novianto, "Pakar Soal Mantan Presiden di Danantara: Tujuan Ekonomi Bergeser Politik," KBR. [Online]. Available: <https://kbr.id/berita/nasional/pakar-soal-mantan-presiden-di-danantara-tujuan-ekonomi-bergeser-politik>
- [8] A. N. Ikhsan, P. Subarkah, A. D. Iftinani, and A. N. Fadilah, "Infotekmesin," *Analisis Sentimen Kenaikan PPN menggunakan Algoritma Naïve Bayes dan Support Vector Machine*, vol. 6, no. 1, Jan. 2025.
- [9] M. Azhari and Parjito, "Analisis Sentimen Opini Publik Program Makan Siang Gratis dengan Random Forest Pada Media X," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, Dec. 2024.
- [10] N. Raisa, N. Riza, and W. I. Rahayu, "JINTEKS (Jurnal Informatika Teknologi dan Sains)," *Analisis sentimen menggunakan svm dan knn pada review drama korea di mydramalist*, vol. 5, no. 4, Nov. 2023.
- [11] P. Arsi and R. Waluyo, "Analisis sentimen wacana pemindahan ibu kota indonesia menggunakan algoritma support vector machine (svm)," *Jurnal Teknologi nformasi dan Ilmu Komputer (JTIIK)*, vol. 8, no. 1, Feb. 2021.
- [12] RevoU, "Apa itu Analisis Sentimen ? Arti, Fungsi, Contoh, FAQs 2025," RevoU. Accessed: Mar. 13, 2025. [Online]. Available: <https://www.revou.co/kosakata/analisis-sentimen>

- [13] E. Yuspita and R. R. Suryono, "Perbandingan Berbagai Metode Klasifikasi Teks Untuk Sentimen Kebijakan Makan Gratis di Indonesia," *The Indonesian Journal of Computer* , vol. 13, no. 5, Oct. 2024, doi: <http://ijcs.net/ijcs/index.php/ijcs/article/view/4440/870>.
- [14] F. Putrawansyah and T. Susanti, "Penerapan Metode Support Vector Machine Terhadap Klasifikasi Jenis Jambu Biji," *Jiko (jurnal informatika dan komputer)*, vol. 8, no. 1, Feb. 2024.
- [15] D. Dahman, "Support vector machine (SVM) - Howdy sysinfo - Medium," *Howdy Sysinfo*, Jul. 12, 2021. Accessed: Mar. 13, 2025. [Online]. Available: <https://medium.com/sysinfo/support-vector-machine-svm-5d95a7d7a547>
- [16] L. Afifah, "Apa itu Confusion Matrix di Machine Learning?," *IlmudataPy*. Accessed: Sep. 28, 2025. [Online]. Available: <https://ilmudatapy.com/apa-itu-confusion-matrix/>
- [17] S. A. Rismawan and Y. Syahidin, "Jurnal Teknik Informatika dan Sistem Informasi ," *Implementasi Website Berita Online Menggunakan Metode Crawling Data Dengan Bahasa Pemrograman Python*, vol. 10, no. 3, Sep. 2023, doi: <https://doi.org/10.35957/jatisi.v10i3.4902>.