ANALISIS SENTIMEN DANANTARA DENGAN METODE *SUPPORT VEKTOR MACHINE* PADA KOMENTAR YOUTUBE

Muhammad Alawi Alatas*[1], Ferdiansyah[2], Rizky Tahara Shita[3], Ika Susanti[4]
Program Studi Teknik Informatika[1,2,3,4], Universitas Budi Luhur[1,2,3,4]
alawialatas1414@gmail.com*[1], ferdiansyah@budiluhur.ac.id[2],
rizky. taharashita@budiluhur.ac.id[3], ika.susanti@budiluhur.ac.id[4]
*Corresponding Author : alawialatas1414@gmail.com

Abstrak

Pembentukan Badan Pengelola Investasi Daya Anagata Nusantara (Danantara) merupakan langkah strategis pemerintah dalam penguatan ekonomi nasional. Beragamnya respons publik mendorong perlunya pendekatan sistematis untuk mengklasifikasikan opini masyarakat. klasifikasi sentimen terhadap Danantara ini menggunakan data komentar *YouTube* dengan algoritma *Support Vector Machine* (SVM). Proses dimulai dari pengumpulan data melalui *YouTube* Data API v3, dilanjutkan dengan pra-pemrosesan teks, pembobotan fitur menggunakan TF-IDF dengan normalisasi L2, dan klasifikasi menggunakan SVM. Pengujian dilakukan dengan tiga variasi rasio data latih dan uji (90:10, 80:20, dan 70:30). Hasil terbaik diperoleh pada rasio 90:10 dengan akurasi 69,49% dan F1-Score berbobot 60,75%. Distribusi kelas yang tidak seimbang, terutama pada kategori Netral, memengaruhi performa model dalam memprediksi kelas minoritas dan komentar yang mengandung sarkasme.

Kata kunci: Analisis Sentimen; Danantara; *Support Vector Machine*

Abstract

The establishment of the Daya Anagata Nusantara Investment Management Agency (Danantara) is a strategic step by the government in strengthening the national economy. The diverse range of public responses necessitates a systematic approach to classifying public opinion. This study focuses on classifying public sentiment toward Danantara using YouTube comment data with the Support Vector Machine (SVM) algorithm. The research process involves data collection via YouTube Data API v3, text preprocessing, feature weighting using TF-IDF with L2 normalization, and SVM-based classification. Testing was conducted using three training-testing split ratios (90:10, 80:20, and 70:30). The best results were achieved at the 90:10 ratio with an accuracy of 69.49% and a weighted F1-Score of 60.75%. Imbalanced class distribution, particularly in the Neutral category, affected the model's ability to predict minority-class and sarcasm-containing comments.

Keyword: *Sentiment Analysis, Danantara, Support Vector Machine*

1. Pendahuluan

Indonesia tengah memasuki era baru dalam tata kelola investasi strategis nasional [1]. Sebagai bagian strategi dari penguatan ekonomi nasional, pemerintah Indonesia membentuk Badan Pengelola Investasi Daya Anagata Nusantara (Danantara). Lembaga ini secara resmi diluncurkan pada 24 Februari 2025 [2]. Tujuan utama nya untuk meningkatkan efisiensi dalam pengelolaan aset strategis milik negara sekaligus menarik investasi internasional guna mempercepat laju Pembangunan ekonomi nasional.

Beragamnya tanggapan publik terhadap terhadap Lembaga BPI Danantara, mulai dari yang setuju, serta ada yang netral, dan ada yang menunjukan ketidakpuasan terhadap kebijakan

tersebut. Namun, dengan jumlah komentar yang melimpah, untuk mengetahui secara garis besar persepsi publik terhadap kebijakan Danantara secara manual menjadi hal yang tidak mungkin. Maka dari itu, perlu dilakukan pendekatan secara sistematis serta efektif dan efisien untuk mengklasifikasikan komentar – komentar tersebut berupa salah satu teknik dari Text Mining yaitu Analisis Sentimen. Analisis Sentimen adalah salah satu teknik yang bertujuan untuk mengidentifikasi dan mengevaluasi polaritas dari suatu Teks. Teknik ini membantu mengidentifikasi dan mengevaluasi sentimen publik terhadap lembaga Danantara, sehingga memberikan gambaran yang lebih jelas berupa klasifikasi opini positif, opini netral, atau opini negatif.

Tujuan penelitian ini adalah mengembangkan sistem analisis sentimen berbasis aplikasi web dengan metode Support Vector Machine (SVM), yang mampu mengklasifikasikan opini masyarakat terhadap lembaga BPI Danantara ke dalam tiga kategori (Positif, Negatif, atau Netral), serta menguji tingkat akurasi metode SVM dalam menganalisis sentimen pada media sosial YouTube.

Penelitian ini diharapkan bermanfaat dengan memberikan evaluasi tingkat akurasi metode Support Vector Machine (SVM) dalam analisis sentimen, serta menjadi referensi bagi akademisi lain dalam menyusun penelitian sejenis di bidang analisis sentimen media sosial.

Analisis sentimen telah menjadi sebuah teknik yang banyak digunakan untuk memahami opini publik terhadap suatu isu atau permasalahan. Penelitian Analisis sentimen terhadap Danantara sudah pernah dilakukan oleh Satria Adi Nugraha (2025) pada platform X. Pada penelitian tersebut metode analisis sentimen yang digunakan adalah Lexicon Based yang memungkinkan klasifikasi sentimen berdasarkan kata dan ekspresi yang digunakan dalam teks dibandingkan kamus yang ada [3]. Studi tersebut mengungkap bahwa persepsi masyarakat terhadap Danantara cenderung beragam, dengan dominasi sentimen negative sebesar 44,9%, sedikit lebih tinggi dibandingkan sentiment positive yang mencapai 44,1%, dan sentiment neutral sebesar 11%. Pendekata dengan Lexicon Based dinilai efisien dan mudah diinterpretasikan, tetapi memiliki kelemahan dalam menangkap konteks yang lebih kompleks seperti sarkasme atau ironi.

Berdasarkan kelemahan dari metode *Lexicon-Based*, peneliti mencoba untuk melakukan pendekatan yang berbeda dari penelitian sebelumnya yaitu dengan menggunakan algoritma *machine learning* untuk meningkatkan akurasi klasifikasi. Dengan fokus pada data YouTube yang memiliki karakteristik teks lebih panjang, kontekstual, dan kompleks dibandingkan X, penelitian ini diharapkan dapat menyajikan hasil klasifikasi yang lebih presisi sehingga memberikan pemahaman yang lebih baik tentang respons/opini publik terhadap Danantara.

2. Kajian Pustaka dan pengembangan hipotesis

2.1.Pengumpulan Data

Tahap awal dimulai dengan pengumpulan data, yaitu proses memperoleh data mentah dari sumber tertentu untuk kemudian dianalisis lebih lanjut. Dalam ranah media sosial, data yang diambil umumnya meliputi isi teks dari unggahan, metadata, serta informasi terkait pengguna. Proses ini memanfaatkan *library* YouTube Data API v3, yang diimplementasikan dalam sebuah skrip Python dan dieksekusi pada lingkungan *Google Colaboratory*. Setelah proses *crawling* selesai, seluruh data komentar yang berhasil diekstraksi disimpan dalam format CSV.

2.2.*Preprocessing*

Preprocessing teks berfungsi mengubah data mentah menjadi lebih terstruktur dan sederhana agar siap digunakan pada tahap analisis penelitian. *Preprocessing* data

memungkinkan proses penambangan berjalan efektif dan efisien [4]. Berikut merupakan tahapan dari Pra-pemrosesan teks:

a. Proses *Case Folding*

   Case Folding, proses mengubah semua kata berhuruf kapital menjadi huruf kecil dapat meningkatkan keterbacaan secara signifikan [5].

b. Proses *Cleansing*

   Cleansing, proses membersihkan teks dokumen dari teks yang tidak dibutuhkan oleh proses seperti, hashtag, html, emoticon, url dan mention [6].

c. Proses *Slangword Normalitation*

   Normalisasi slang word menyamakan istilah tidak baku seperti singkatan atau bahasa gaul ke bentuk standar bahasa Indonesia.

d. Proses *Tokenizing*

   Tokenisasi adalah proses mendasar dalam analisis teks yang melibatkan penguraian teks tertentu menjadi unit-unit yang lebih kecil yang dikenal sebagai token [7].

e. Proses *Stopword Removal*

   Stopword Removal merujuk pada kata-kata yang sering muncul dan sering tidak penting dalam teks tertentu yang biasanya dihilangkan selama pemrosesan data [7].

f. Proses *Stemming*

   Stemming adalah proses linguistik yang mereduksi kata-kata ke bentuk dasarnya [7].

2.3.Pelabelan Data

Pelabelan adalah fase penting dalam pengolahan data teks untuk pembelajaran berbasis pengawasan (*supervised learning*), di mana setiap entri diberikan label atau kategori berdasarkan karakteristiknya. Misalnya, dalam analisis sentimen, teks dari komentar atau ulasan diberi tanda sebagai positif, negatif, atau netral agar model dapat belajar mengenali pola-pola emosional [8]

2.4.Transformasi dan Pembobotan

Merupakan proses perhitungan atau pengekstrakan kata menjadi sebuah angka berbentuk vektor yang digunakan untuk menentukan bobot dari sebuah kata dalam sebuah dokumen atau korpus [9]. Fitur seleksi merupakan proses mengubah data kategorikal menjadi data numerik [10]. Hasil penghitungan TF-IDF menghasilkan representasi vektor teks yang selanjutnya dapat digunakan dalam algoritma seperti pembelajaran mesin atau deteksi kemiripan teks.

$$\text{TF} = \frac{Jumlah\ kemunculan\ kata\ t\ dalam\ dokumen\ d}{Total\ jumlah\ kata\ dalam\ dokumen\ d} \quad (1)$$

$$IDF = ln\left(\frac{N+1}{DF+1}\right) + 1 \quad (2)$$

$$tfidf(t,d,D) = tf(t,d) * idf(t,D) \quad (3)$$

$$\frac{v}{\|v_2\|} = \frac{v}{\sqrt{v_1^2 + v_2^2 \dots\dots\dots + v_n^2}} \quad (4)$$

2.5.Klasifikasi *Support Vector Machine*

SVM adalah kependekan dari Support Vector Machine, sebuah algoritma yang membagi dua kelompok kelas data menggunakan fungsi linear dalam ruang fitur yang memiliki dimensi yang tinggi [11]. Prinsip kerja fundamental dari SVM adalah menemukan sebuah bidang pemisah atau *hyperplane* yang paling optimal untuk memisahkan data ke dalam kelas-kelas yang berbeda di dalam ruang fitur (*feature space*).

a. Fungsi Keputusan

Konsep paling dasar dari SVM adalah kemampuannya untuk menemukan sebuah garis atau bidang pemisah, yang secara teknis disebut *hyperplane*, untuk memisahkan data ke dalam kelas-kelas yang berbeda.

$$f(x) = w \cdot x + b \tag{5}$$

Keterangan:

- f(x): Merupakan skor prediksi untuk sebuah data input x.
- w: Adalah vektor bobot (*weights*), yang merupakan sekumpulan nilai yang dipelajari oleh model selama pelatihan. Vektor ini menentukan orientasi atau kemiringan dari *hyperplane*.
- x: Adalah vektor fitur dari data input
- b: Adalah *bias*, sebuah nilai skalar yang berfungsi untuk menggeser posisi *hyperplane* tanpa mengubah kemiringannya.

b. Kondisi Klasifikasi

SVM tidak hanya bertujuan untuk mengklasifikasikan data dengan benar, tetapi juga untuk memaksimalkan jarak atau *margin* antara *hyperplane* dengan titik data terdekat dari setiap kelas.

$$y \cdot (w \cdot x + b) \geq 1 \tag{6}$$

Keterangan

- y: Merupakan label kelas asli dari data, yang harus bernilai +1 untuk kelas target dan -1 untuk kelas lainnya.
- Logika**:** Jika hasil perkalian ini lebih besar atau sama dengan 1, artinya data tidak hanya berada di sisi yang benar dari *hyperplane*, tetapi juga berada di luar zona *margin*. Ini dianggap sebagai klasifikasi yang sudah baik. Jika hasilnya kurang dari 1, ini dianggap sebagai "kesalahan" yang harus diperbaiki.

c. Stochastic Gradient Descent (SGD)

Untuk menemukan nilai w dan b yang optimal, model perlu melalui proses "belajar" secara iteratif. Algoritma yang digunakan untuk proses pembelajaran ini adalah *Stochastic Gradient Descent* (SGD).

Rumus Pembaruan Bobot $$w \leftarrow w - \eta \cdot (2\lambda w - y \cdot x) \tag{7}$$

Rumus Pembaruan Bias $$b \leftarrow b + \eta \cdot y \tag{8}$$

Keterangan:

- η (eta): learning_rate.
- λ (lambda): lambda_param

2.6.*Confussion Matrix*

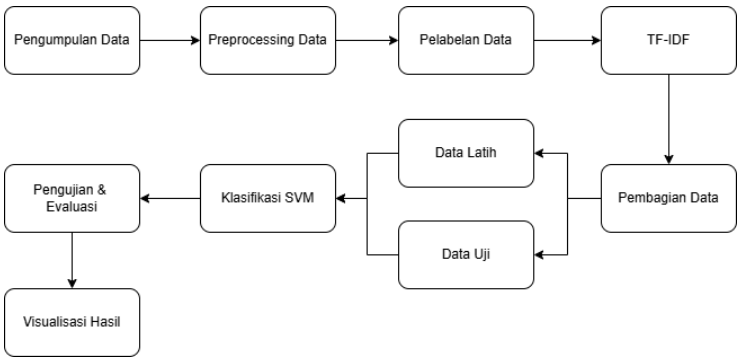
Confusion matrix adalah sebuah metode evaluasi performa model klasifikasi yang menyajikan rincian prediksi benar dan salah dalam format tabel yang jelas. Untuk masalah dengan banyak kelas (multikelas), setiap sel dalam matriks menunjukkan hubungan spesifik antara nilai aktual dan nilai prediksi. Prediksi yang benar diidentifikasi sebagai True Positive (TP) untuk kelas tersebut, sedangkan setiap kesalahan klasifikasi dianalisis dari dua sudut pandang, sebagai False Negative (FN) untuk kelas yang sebenarnya (aktual), dan False Positive (FP) untuk kelas yang salah diprediksi.

Confussion Matrix		**Prediksi**		
		Positif	Negatif	Netral
Aktual	Positif	TP (Positif)	FN (Positif) & FP (Negatif)	FN (Positif) & FP(Netral)
	Negatif	FN (Negatif) & FP (Positif)	TP (Negatif)	FN (Negatif) & FP (Netral)
	Netral	FN (Netral) & FP (Positif)	FN (Netral) & FP (Negatif)	TP (Netral)

Gambar 1 *Confussion Matrix* 3 x 3

3. Metode Penelitian

Dalam penelitian ini metode penelitian pada klasifikasi analisis sentimen memiliki beberapa tahap sistematis yang dilakukan pada penelitian ini, dimulai dari pengumpulan data hingga visualisasi hasil. Untuk memberikan gambaran yang jelas, alur penelitian ini divisualisasikan dalam bentuk *flowchart* pada Gambar 2 berikut.



Gambar 2 *flowchart* metode penelitian

Tahapan pertama dalam penelitian ini adalah pengumpulan data yang dilakukan dengan teknik *Crawling* menggunakan *Google Colab,* dimana data yang berhasil diperoleh disimpan dalam file berformat *Comma Separated Values*. Lalu tahapan selanjutnya adalah *preprocessing* dimana data yang masih mentah, hasil dari *Crawling* dibersihkan dengan berbagai tahapan, agar bisa diproses. Lalu tahapan selanjutnya adalah Pelabelan yang dilakukan oleh ahli pakar. Lalu tahapan selanjutnya adalah transformasi teks dan pembobotan dengan teknik *term frequency-inverse document frequency* agar data yang sudah dibersihkan pada tahapan sebelumnya, dapat diberikan bobot dan dapat dipelajari oleh model. Lalu tahapan selanjutnya adalah pembagian data dimana data dibagi menjadi dua bagian yaitu data latih dan data uji. Lalu tahapan selanjut nya adalah klasifikasi model *Support Vector Machine* untuk analisis sentimen. Ditahap selanjutnya adalah proses Evaluasi untuk menganalisa hasil dari proses klasifikasi. Dan tahapan

terakhir adalah visualisasi untuk membuat sebuah visual dari hasil dari pengujian dan evaluasi yang sudah dilakukan.

4. Hasil dan Pembahasan

4.1.Pengumpulan Data

Tahap awal dimulai dengan pengumpulan data, yaitu proses memperoleh data mentah dari sumber tertentu untuk kemudian dianalisis lebih lanjut. Dalam ranah media sosial, data yang diambil umumnya meliputi isi teks dari unggahan, metadata, serta informasi terkait pengguna. Proses ini memanfaatkan *library* YouTube Data API v3, yang diimplementasikan dalam sebuah skrip Python dan dieksekusi pada lingkungan *Google Colaboratory*. Setelah proses *crawling* selesai, seluruh data komentar yang berhasil diekstraksi disimpan dalam format CSV.

4.2.Preprocessing

Pada *preprocessing* dilakukan serangkaian tahapan yang mencakup beberapa langkah, antara lain *case folding*, pembersihan data (*cleaning*), tokenisasi, normalisasi slangword, penghapusan stopword, dan stemming. Tujuan dari tahapan ini adalah untuk mengubah data mentah yang masih belum terstruktur menjadi informasi yang lebih rapi dan siap untuk dianalisis. Contoh hasil dari proses ini bisa dilihat pada Tabel 1 berikut.

Tabel 1 Hasil dari *preprocessing*

Tahapan	*Hasil*
Teks Asli	🤔 Positifnya Masyarakat Belajar MENGELOLA INVESTASI, Negatifnya kalau ADA YG MENYELEWENGKAN DANA.. 🤔
Case Folding	🤔 positifnya masyarakat belajar mengelola investasi, negatifnya kalau ada yg menyelewengkan dana.. 🤔
Cleansing	positifnya masyarakat belajar mengelola investasi negatifnya kalau ada yg menyelewengkan dana
Tokenizing	['positifnya', 'masyarakat', 'belajar', 'mengelola', 'investasi', 'negatifnya', 'kalau', 'ada', 'yg', 'menyelewengkan', 'dana']
Slang Word Normalization	['positifnya', 'masyarakat', 'belajar', 'mengelola', 'investasi', 'negatifnya', 'kalau', 'ada', 'yang', 'menyelewengkan', 'dana']
Stopword Removal	['positifnya', 'masyarakat', 'belajar', 'mengelola', 'investasi', 'negatifnya', 'menyelewengkan', 'dana']
Stemming	positif masyarakat ajar kelola investasi negatif seleweng dana

4.3.Pelabelan Data

Setelah melalui tahap 2 *preprocessing* teks, data komentar dilabeli secara manual oleh dosen ahli pakar untuk menjamin objektivitas dan validitas. Hasilnya adalah dataset final berlabel yang digunakan sebagai dasar pembobotan fitur dan pelatihan model, dengan distribusi label ditunjukkan pada Tabel 2.

Tabel 2 Distribusi label oleh ahli pakar

Total Data	*Label Positif*	*Label Negatif*	*Label Netral*
1176	327	722	127

4.4.Transformasi dan Pembobotan

Setelah data teks berhasil dibersihkan melalui tahap *preprocessing,* tahap selanjutnya adalah pembobotan/transformasi data sehingga data dapat diproses oleh algoritma *machine learning.* Pembobotan kata melibatkan perhitngan *Term-Frequency* (TF), *Inverse Document Frequency* (IDF)*, Term Frequency – Inverse Document Frequency* (TF-IDF), serta Normalisasi L2. Berikut adalah sampel data sebagai simulasi proses perhitungan TF-IDF dengan Normalisasi L2 bisa dilihat pada Tabel 3.

Tabel 3 Sampel data untuk perhitungan TF-IDF Norm L2

Doc	*Text*	*Label*
D1	benar semua program setuju bagus bikin jelek koruptor cuma waspada sama koruptor	Positif
D2	semoga danantara jadi badan investasi indonesia global tingkat dunia	Positif
D3	selama pelaku koruptor tidak hukum mati ragu	Netral
D4	kalau danantara gagal siap indonesia hancur	Negatif

Berdasarkan rumus 1 tentang perhitungan TF, lalu rumus 2 tentang perhitungan IDF, serta rumus 3 tentang perhitungan TF-IDF dan rumus 4 tentang Normalisasi L2 diperoleh hasil dan dapat dilihat sampel hasil dari perhitungan tersebut pada tabel 4 berikut.

Tabel 4 Contoh hasil perhitungan proses TF-IDF Norm L2

Term	*TF*	*IDF*	*TF-IDF*	*Norm L2*
bagus	0,08333	1,916291	0,159691	0,28299703
benar	0,08333	1,916291	0,159691	0,28299703
bikin	0,08333	1,916291	0,159691	0,28299703
cuma	0,08333	1,916291	0,159691	0,28299703

4.5.Pengujian

Setelah proses pembobotan dengan teknik *Term-Frequency Inverse Document Frequency*, dataset dibagi menjadi dua bagian terpisah untuk keperluan pelatihan dan pengujian model. Proses ini menggunakan metode *Random Sampling* dengan 3 rasio perbandingan data latih dan data uji yaitu 90:10, 80:20, dan 70:30. Berikut adalah hasil dari pengujian dengan berbagar rasion beserta Confussion Matrix dan perhitungan pada masing – masing kelas. Pada Tabel 5 hasil Confussion Matrix rasio 90:10.

Tabel 5 Confussion Matrix untuk rasio 90:10

Confussio Matrix	*Prediksi Positif*	*Prediksi Negatif*	*Prediksi Netral*	*Jumlah*
Aktual Positif	7	21	0	28
Aktual Negatif	0	75	0	75
Aktuan Netral	0	15	0	15
Jumlah	7	111	0	118

Berdasarkan Tabel 5 dan rumus 9, 10, 11 dan 12 dapat dihitung akurasi, presisi, recall dan f1-score. Pada Tabel 6 bisa dilihat untuk hasil perhitungan akurasi, presisi, recall dan f1-score untuk rasio 90:10.

Tabel 6 Perhitungan Metrik evaluasi rasio 90:10

Kelas	*Positif*	*Negatif*	*Netral*
Presisi	100%	67,57%	0%
Recall	25%	100%	0%
F1-Score	40%	80,65%	0%
Akurasi		69,49%	

Pada Tabel 7 bisa dilihat untuk hasil perhitungan akurasi, presisi, recall dan f1-score untuk rasio 80:20.

Tabel 7 Confussion Matrix untuk rasio 80:20

Confussio Matrix	*Prediksi Positif*	*Prediksi Negatif*	*Prediksi Netral*	*Jumlah*
Aktual Positif	13	47	0	60
Aktual Negatif	1	150	0	151
Aktuan Netral	1	24	0	25
Jumlah	15	221	0	236

Pada Tabel 8 dapat dilihat hasil *confussion matrix* yang telah diperoleh, maka dapat dilakukan proses perhitungan untuk evaluasi model. Berikut disajikan hasil dari perhitungan tersebut.

Tabel 8 Perhitungan Metrik evaluasi rasio 80:20

Kelas	*Positif*	*Negatif*	*Netral*
Presisi	86,67%	67,87%	0%
Recall	21,67%	99,34%	0%
F1-Score	34,67%	80,65%	0%
Akurasi		69,07%	

Pada Tabel 9 bisa dilihat untuk hasil perhitungan akurasi, presisi, recall dan f1-score untuk rasio 70:30.

Tabel 9 Confussion Matrix untuk rasio 70:30

Confussio Matrix	Prediksi Positif	Prediksi Negatif	Prediksi Netral	Jumlah
Aktual Positif	17	67	0	84
Aktual Negatif	3	226	0	229
Aktuan Netral	1	39	0	40
Jumlah	21	332	0	353

Pada Tabel 10 dapat dilihat hasil confusion matrix yang telah diperoleh, maka dapat dilakukan proses perhitungan untuk evaluasi model. Berikut disajikan hasil dari perhitungan tersebut.

Tabel 10 Perhitungan Metrik evaluasi rasio 70:30

Akurasi	68,84%		
Kelas	*Positif*	*Negatif*	*Netral*
Presisi	80,95%	68,07%	0%
Recall	20,24%	98,69%	0%
F1-Score	32,38%	80,57%	0%
Akurasi		68,84%	

4.6.Analisa Pengujian

Pada penelitian ini, dilakukan evaluasi performa model *Support Vector Machine* (SVM) untuk klasifikasi sentimen terkait peluncuran Danantara. Pengujian dilakukan dengan menerapkan tiga skenario rasio pembagian data yang berbeda (90:10, 80:20, dan 70:30) untuk melihat pengaruh jumlah data latih terhadap kinerja model. Performa model kemudian diukur menggunakan serangkaian metrik evaluasi, yang meliputi akurasi, presisi, recall, dan F1-score. Hasil pengujian bisa dilihat pada Tabel 11 berikut

Tabel 11 Analisa Hasil Pengujian

No	*Rasio*	*Akurasi*	*Presisi*	*Recall*	*F1-Score*	*Prediksi Benar*	*Jumlah Data*
1	90:10	69,49%	66,67%	69,49%	60,75%	82	118
2	80:20	69,07%	65,46%	69,07%	60,41%	163	236
3	70:30	68,84%	63,42%	68,84%	59,97%	243	353

Hasil pengujian menunjukkan bahwa rasio 90% data latih dan 10% data uji menghasilkan performa model paling optimal dengan Akurasi 69,49% dan F1-Score 60,75%, yang membuktikan semakin banyak data latih digunakan, semakin baik model mengenali pola. Namun, analisis Confusion Matrix mengungkap kecenderungan model untuk memprediksi kelas mayoritas (Negatif) dan gagal mengenali kelas minoritas (Netral) di semua skenario pengujian

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan keseluruhan proses pada penelitian ini, dapat ditarik kesimpulan bahwa penelitian ini berhasil mengembangkan sistem analisis sentimen berbasis web menggunakan metode Support Vector Machine (SVM) yang mencakup pengelolaan dataset, pra-pemrosesan data, pelatihan model, dan prediksi sentimen. Sistem dapat mengklasifikasikan opini publik tentang BPI Danantara ke dalam kategori Positif, Negatif, dan Netral, namun cenderung bias ke kelas mayoritas (Negatif) akibat distribusi data yang tidak seimbang. Pengujian dengan berbagai rasio data menunjukkan performa terbaik pada pembagian 90% data latih dan 10% data uji, dengan akurasi 69,49% dan F1-Score tertimbang 60,75%, meskipun model masih kesulitan mengenali kelas minoritas (Netral).

5.2 Saran

Penelitian ini masih memiliki keterbatasan, sehingga pengembangan ke depan disarankan untuk mengatasi ketidakseimbangan data dengan teknik seperti oversampling

(SMOTE) atau undersampling, mengeksplorasi model alternatif berbasis deep learning seperti LSTM, BiLSTM, atau Transformer (BERT) untuk pemahaman konteks yang lebih baik, serta mengembangkan representasi fitur selain TF-IDF, seperti *word embeddings* (*Word2Vec* atau GloVe) agar model lebih mampu menangkap hubungan semantik antar kata.

Referensi

[1] A. M. M. Nasoha, "DANANTARA: Reformasi Investasi Strategis dalam Hukum dan Ekonomi -Fakultas Syariah UIN Surakarta." Accessed: Jul. 15, 2025. [Online]. Available: https://syariah.uinsaid.ac.id/danantara-reformasi-investasi-strategis-dalam-hukum-dan-ekonomi/

[2] Putra Ambisius, "Prabowo Resmikan Wisma Danantara, Pusat Pengelolaan Investasi Negara Triliunan Dolar." Accessed: Jul. 15, 2025. [Online]. Available: https://news.ambisius.com/2025/07/02/politik/prabowo-resmikan-wisma-danantara-pusat-pengelolaan-investasi-negara-triliunan-dolar

[3] S. Adi Nugraha, "Penerapan Lexicon Based Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Danantara," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4949–4957, 2025, doi: 10.36040/jati.v9i3.13836.

[4] I. Kurniawan, A. Lia Hananto, S. Shofia Hilabi, A. Hananto, B. Priyatna, and A. Yuniar Rahman, "Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 731–740, 2023, [Online]. Available: http://jurnal.mdp.ac.id

[5] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.

[6] T. Muhayat, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Komentar Video Youtube Menggunakan Support Vector Machines," *Progresif J. Ilm. Komput.*, vol. 19, no. 1, p. 231, 2023, doi: 10.35889/progresif.v19i1.1060.

[7] M. L. Hermanto, O. Ardhillah, and A. Ibrahim, "Danantara di Platform X dengan Metode SVM," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 4, pp. 6779–6785, 2025.

[8] S. Biswas, K. Young, and J. Griffith, "A Comparison of Automatic Labelling Approaches for Sentiment Analysis," pp. 312–319, 2022, doi: 10.5220/0011265900003269.

[9] K. Tri Putra, M. Amin Hariyadi, and C. Crysdian, "Perbandingan Feature Extraction Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis Svm," *J. Cahaya MAndalika*, p. 1449, 2023.

[10] T. M. Permata Aulia, N. Arifin, and R. Mayasari, "Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19," *SINTECH (Science Inf. Technol. J.*, vol. 4, no. 2, pp. 139–145, 2021, doi: 10.31598/sintechjournal.v4i2.762.

[11] L. Rofiqi and M. Akbar, "Analisis Sentimen Terkait RUU Perampasan Aset dengan Support Vector Machine," *JEKIN - J. Tek. Inform.*, vol. 4, no. 3, pp. 529–538, 2024, doi: 10.58794/jekin.v4i3.824.