

## Perbandingan *Naïve Bayes* dan SVM untuk Analisis Sentimen Pengguna Aplikasi X Terhadap Danantara

DOI: <http://dx.doi.org/10.35889/jutisi.v14i3.3337>

Creative Commons License 4.0 (CC BY – NC)

Aulia Riviani\*, Cahyono Budy Santoso<sup>2</sup>

Sistem Informasi, Universitas Pembangunan Jaya, Tangerang Selatan, Indonesia

\*e-mail Corresponding Author: aulia.riviani@student.upj.ac.id

### Abstract

*The Indonesian Investment Management Agency Danantara program is an initiative by the Indonesian government aimed at strengthening the national economy through the management of strategic assets and the enhancement of global investment. The launch of this program has generated various public responses on social media, particularly on the X (Twitter) platform. This study aims to analyze public sentiment in Indonesia toward the Danantara program and to compare the performance of two machine learning algorithms, namely Naïve Bayes and Support Vector Machine (SVM). The analysis process begins with data collection through web crawling, followed by data preprocessing, sentiment labeling using a hybrid lexicon-based approach, TF-IDF weighting, and classification using both algorithms. Model Evaluation is conducted using accuracy, precision, recall, and F1-score metrics. The results indicate that the SVM algorithm achieves an accuracy of 87% while the Naïve Bayes algorithm attains an accuracy of 83,7%. These findings demonstrate that SVM outperforms Naïve Bayes in sentiment text classification. The result of this study is expected to serve as a reference for understanding public perception and as a consideration for policymakers.*

**Keywords:** Sentiment Analysis; Danantara; Support Vector Machine; Naïve Bayes; Lexicon-Based

### Abstrak

Program Badan Pengelola Investasi Danantara merupakan inisiatif pemerintah Indonesia yang bertujuan untuk memperkuat perekonomian nasional melalui pengelolaan aset strategis dan peningkatan investasi global. Peluncuran program ini memunculkan berbagai tanggapan di media sosial, terutama di platform X (*Twitter*). Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap program Danantara serta membandingkan kinerja dua algoritma machine learning, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM). Proses analisis dimulai dengan pengumpulan data dengan crawling, preprocessing data, pelabelan menggunakan pendekatan *hybrid Lexicon-based*, pembobotan TF-IDF, serta klasifikasi menggunakan kedua algoritma. Evaluasi dilakukan dengan metrik accuracy, precision, recall, dan f1-score. Hasil penelitian menunjukkan bahwa algoritma SVM memperoleh nilai akurasi sebesar 87% dan *Naïve Bayes* memperoleh akurasi sebesar 83,7%. Temuan ini menunjukkan bahwa SVM lebih unggul dalam mengklasifikasikan teks sentimen. Hasil penelitian ini diharapkan dapat menjadi referensi dalam memahami persepsi publik serta menjadi bahan pertimbangan bagi pemangku kebijakan.

**Kata kunci:** Analisis Sentimen; Danantara; Support Vector Machine; Naïve Bayes; Lexicon-Based

### 1. Pendahuluan

Badan Pengelola Investasi Daya Anagata Nusantara (Danantara) merupakan program terbaru milik pemerintah Indonesia yang telah diresmikan pada 24 Februari 2025[1]. Pembentukan Danantara menjadi salah satu upaya dalam memperkuat perekonomian nasional. Badan pengelola investasi ini diharapkan mampu meningkatkan efisiensi pengelolaan aset strategis nasional serta menarik investasi global guna mempercepat pertumbuhan ekonomi

Indonesia[2][3]. Di tengah kondisi ekonomi yang terus berubah, peluncuran program Danantara dipandang sebagai katalisator dalam mendorong pembangunan berkelanjutan.

Meskipun peluncuran Danantara membawa harapan bagi masa depan perekonomian Indonesia, program ini memicu berbagai macam tanggapan masyarakat Indonesia, termasuk di media sosial. X (Sebelumnya *Twitter*) menjadi media sosial yang dijadikan sebagai tempat untuk berinteraksi dan menyampaikan opini[4][5][6]. Tercatat ada sekitar 24 juta pengguna aktif di Indonesia[7]. Sebagai sosial media yang bersifat terbuka serta proses penyebaran informasi yang begitu cepat dan *real-time* menjadikan X sebagai salah satu sumber data yang sangat relevan untuk memetakan persepsi publik. Referensi [8] menyatakan bahwa media sosial dianggap mampu dan lebih dapat diandalkan untuk mengumpulkan opini masyarakat dibandingkan survei.

Memahami persepsi publik menjadi penting untuk isu strategis seperti Danantara. Kebijakan dan keputusan investasi yang berdampak luas pada layanan publik, kinerja BUMN, dan iklim investasi memerlukan dukungan sosial yang memadai. Pemetaan sentimen berbasis data dari media sosial dapat berfungsi sebagai alat peringatan dini bagi pengambil kebijakan untuk mengidentifikasi topik sensitif, memperbaiki miskonsepsi, serta merancang strategi komunikasi yang lebih tepat sasaran, terutama pada fase awal implementasi program yang rentan terhadap disinformasi. Oleh karena itu, analisis sentimen menjadi penting untuk mengetahui bagaimana persepsi masyarakat Indonesia terhadap program BPI Danantara.

Analisis sentimen menjadi salah satu pendekatan dalam memahami persepsi publik terhadap berbagai bidang, seperti kesehatan, isu politik, layanan publik, fenomena sosial, hingga kebijakan pemerintah maupun isu nasional[9][10][11][12][13]. Terdapat berbagai metode yang biasa digunakan dalam analisis sentimen, salah satunya adalah *machine learning* yang bisa dilakukan dengan algoritma *Naïve Bayes*, *Random Forest*, atau *Support Vector Machine*.

Beberapa penelitian telah dilakukan dengan membandingkan kinerja algoritma SVM dan *Naïve Bayes* dalam analisis sentimen pada berbagai konteks isu publik hingga lembaga keuangan maupun investasi dan menunjukkan bahwa SVM cenderung menghasilkan performa yang lebih stabil pada dan memiliki akurasi lebih unggul[14][15][16]. Selain itu, penelitian yang dilakukan Nugraha [2] dengan menggunakan 5337 data mentah pada topik Danantara menghasilkan fakta sentimen publik mengenai Danantara cenderung negatif (44,9%), tetapi pendekatannya murni *lexicon-based* tanpa adanya pendekatan *machine learning*.

Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan pendekatan *hybrid* yang mengombinasikan pelabelan menggunakan *lexicon* dengan *machine learning*. Proses pelabelan awal dilakukan secara otomatis menggunakan dua kamus sentimen Bahasa Indonesia yaitu InSet Lexicon yang dikembangkan oleh Fajri dan Gemala[17] serta SentiStrength\_id yang dikembangkan oleh Devid Haryalesmana[18]. Kemudian hasil dari pelabelan akan digunakan sebagai data latih bagi dua algoritma *classifier*, yaitu *Support Vector Machine* dan *Naïve Bayes*.

Tujuan dari penelitian ini adalah untuk membandingkan kinerja algoritma *Support Vector Machine* dan *Naïve Bayes* dalam mengklasifikasikan sentimen pengguna X Indonesia terhadap program Danantara. Perbandingan dilakukan berdasarkan nilai evaluasi kinerja berupa akurasi untuk menentukan algoritma yang memiliki performa terbaik dalam mengklasifikasikan sentimen. Selain itu, penelitian ini bertujuan untuk mengidentifikasi kecenderungan sentimen pengguna aplikasi X terhadap Danantara. Hasil dari penelitian ini diharapkan dapat menjadi sumber informasi pendukung dalam memahami kecenderungan sentimen masyarakat di media sosial aplikasi X terhadap Danantara serta menjadi bahan pertimbangan bagi pemangku kebijakan.

## 2. Tinjauan Pustaka

Penelitian mengenai analisis sentimen terhadap program Danantara telah dilakukan oleh Nugraha (2025) dengan judul *Penerapan Lexicon Based Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Danantara*. Penelitian tersebut menggunakan metode *lexicon-based* yang dapat memberikan wawasan bagi pemangku kepentingan menunjukan bahwa dari 5337 data *tweet* mentah yang berkaitan dengan Danantara. Hasilnya menunjukkan bahwa 44.9% negatif, 44.1% positif, dan 11% netral[2]. Penelitian ini memberikan gambaran umum persepsi publik, namun belum menyertakan tahapan klasifikasi berbasis *machine learning* sehingga pemetaan sentimen masih bergantung sepenuhnya pada skor dari kamus *lexicon*.

Selain itu, ada studi lain [19] yang telah menerapkan pendekatan *machine learning* pada topik Danantara menggunakan SVM dengan total dataset sebanyak 1175 tweet dengan memanfaatkan pelabelan otomatis yang menghasilkan tingkat akurasi sebesar 79,57%, *precision* 79,54%, *recall* 69,62%, dan *F1-Score* 71,63%. Penelitian ini menunjukkan pendekatan *supervised* seperti SVM dapat memberikan performa klasifikasi yang terukur melalui metrik akurasi dan *F1-Score* serta. Berdasarkan literatur yang ada, literatur mengenai Danantara masih terbatas dan belum banyak yang membahas pelabelan *hybrid lexicon* sebagai dasar latih untuk membandingkan model klasifikasi.

Kurnia [20] melakukan penelitian analisis sentimen terhadap Bank BTN dengan menerapkan kombinasi metode *Naïve Bayes*, SMOTE, dan Adaboost. Hasil penelitian menunjukkan bahwa *Naïve Bayes* menghasilkan akurasi sebesar 75% yang kemudian meningkat menjadi 85,87% ketika digabungkan dengan SMOTE. Temuan ini memperlihatkan bahwa penyeimbangan data seperti SMOTE dapat meningkatkan performa algoritma klasifikasi, terutama pada dataset yang mengalami ketidakseimbangan kelas.

Berbeda dengan kajian sebelumnya, penelitian ini secara khusus berfokus pada analisis sentimen terhadap Danantara dengan membandingkan kinerja dua algoritma *machine learning*, yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes* dengan metode pelabelan *hybrid* menggunakan InSet Lexicon dan SentiStrength\_id untuk meningkatkan reliabilitas label dengan saling melengkapi. InSet berfungsi memberikan polaritas dasar, sedangkan SentiStrength\_id menangkap intensitas emosi. Hasil dari pelabelan data opini dikelompokkan ke dalam dua kelas sentimen, yaitu positif dan negatif sehingga hasil klasifikasi lebih fokus dalam menggambarkan kecenderungan respon masyarakat serta menambahkan Teknik SMOTE untuk mengatasi ketidakseimbangan kelas. Evaluasi performa dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F-1 Score*.

### 3. Metodologi

Penelitian ini dilakukan dengan membandingkan kinerja algoritma *Naïve Bayes* dan *Support Vector Machine* dalam analisis sentimen publik terhadap program Danantara. Data yang digunakan merupakan data sekunder yang didapat melalui proses *crawling* dengan *library* Python TweetHarvest. Metode penelitian yang digunakan adalah kombinasi *Natural Language Processing* pada tahap *pre-processing* data dengan metode *Machine Learning* dalam tahapan klasifikasi sentimen dengan algoritma *Naïve Bayes* dan *Support Vector Machine*. Selanjutnya, kinerja model dievaluasi dengan *confusion matrix*.



Gambar 1. Alur Penelitian

Alur Penelitian yang dilakukan dalam penelitian ini mencakup beberapa tahap seperti yang dilampirkan pada gambar 1, yaitu sebagai berikut:

#### 1) **Crawling Data**

Data dikumpulkan melalui *scrapping* X yang melibatkan pengumpulan data secara otomatis dengan menggunakan *library* Python TweetHarvest. Proses *crawling* dilakukan pada bulan April 2025 dengan rentang 25 Februari - 14 April 2025 dan disimpan dalam format CSV.

#### 2) **Pre-processing Data**

Data mentah yang telah diperoleh dari *crawling* akan melalui proses selanjutnya yaitu, pra-pemrosesan teks yang meliputi *cleaning*, *tokenization*, *normalization*, *stopwords removal*, dan *stemming*. Tujuan dari tahap ini adalah untuk membersihkan teks agar siap untuk dilakukan analisis oleh algoritma SVM dan *Naïve Bayes*.

##### a) **Cleaning Data**

Pada tahap *cleaning* data, elemen-elemen yang tidak dibutuhkan seperti URL, mention (@username), tanda tagar, emoji, tanda baca, spasi ganda akan dihapus. Kemudian dilakukan *casefolding*, yaitu proses seluruh huruf menjadi *lowercase* agar lebih konsisten dan menghapus tweet yang duplikat.

- b) Tokenization**  
Tokenisasi adalah proses memecah kalimat atau teks berdasarkan kata penyusunnya[21]. Proses ini membagi teks menjadi unit-unit terkecil yang disebut dengan token.
- c) Normalization**  
Normalisasi dilakukan untuk menyamakan berbagai variasi kata ke bentuk yang bermakna sama [22], yaitu mengubah kata informal seperti “gk”, “nggak”, “ga” menjadi kata formal “tidak”.
- d) Stopwords Removal**  
Kata-kata umum yang sering muncul dan tidak memberikan informasi penting seperti “yang”, “di”, “dan”, “ke” dihapus dari teks [2]. Penghapusan *stopwords* membantu memfokuskan analisis pada kata-kata yang memiliki makna lebih penting. Pada tahap ini dibuat negasi stopwords seperti kata “tidak”, “bukan”, “jangan”, “tak”, “tidak”, “belum”, “enggak” guna mempertahankan nilai negatif.
- e) Pelabelan**  
Tahap ini akan mengelompokkan sentimen menggunakan metode *Lexicon-based* dengan menggabungkan kamus InSet Lexicon dan SentiStrength\_id. Kosakata di dalam kedua kamus tersebut memiliki bobot -5 sampai dengan 5 yang kemudian distandarisasi menjadi dua kategori, yaitu negatif (-1) dan positif (+1). Total terdapat 3.913 kata positif dan 7.263 kata negatif.
- f) Stemming**  
Pada tahap ini terjadi proses mengubah kata ke bentuk dasarnya [21]. *Stemming* dilakukan dengan menggunakan *library* Sastrawi.
- 3) Splitting Data**  
Pada tahap ini, dataset yang telah melewati proses *filtering* akan dibagi menjadi data pelatihan dan pengujian. Data yang dilatih sebanyak 80% dan data yang akan diuji sebanyak 20%.
- 4) Pembobotan TF-IDF**  
Setelah data latih tersedia, diterapkan pembobotan kata dengan TF-IDF untuk menentukan seberapa penting suatu kata dan frekuensi kemunculannya dalam dokumen, sehingga teks bisa diubah menjadi data numerik yang dapat diproses oleh kedua algoritma [9]. Pada tahap ini ditemukan bahwa distribusi kelas sentimen tidak seimbang yang mana label sentimen negatif mendominasi dibandingkan positif. Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan SMOTE (*Synthetic Minority Oversampling Technique*) pada data pelatihan. SMOTE bekerja dengan membuat data sintesis untuk kelas minoritas berdasarkan jarak antar tetangga terdekat sehingga distribusi menjadi lebih seimbang[23].
- 5) Klasifikasi Data**  
Sebelum melakukan klasifikasi akhir dengan kedua algoritma, data akan divalidasi dengan Teknik *5-fold Cross Validation* sebagai proses evaluasi awal terhadap masing-masing model, baik pada kondisi *baseline* dan setelah penerapan SMOTE. *Fold Cross Validation* digunakan untuk membandingkan efektivitas keseimbangan data dan memberikan gambaran performa yang lebih stabil melalui metrik *F1-Weighted*. Hasil *cross validation* digunakan untuk menentukan apakah SMOTE memberikan peningkatan yang signifikan. Oleh karena itu, hanya model dengan hasil terbaik pada *cross validation* yang dilanjutkan pada proses klasifikasi akhir menggunakan data uji (20%). Pada bagian klasifikasi akhir, pemodelan data uji dilakukan dengan menggunakan *Support Vector Machine* dan *Naïve Bayes*. Kedua algoritma ini dipilih karena banyak digunakan dalam penelitian analisis sentimen berbasis media sosial.
- a) Klasifikasi dengan SVM  
*Support Vector Machine* bekerja dengan mencari *hyperplane* yang dapat memisahkan data ke dalam kelas sentimen positif dan negatif. Pada penelitian ini, proses klasifikasi sentimen dengan SVM dilakukan menggunakan kernel linear.

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i(w^T x_i + b)) \quad \dots\dots\dots (1)$$

Keterangan:

- $w$  : Bobot model  
 $b$  : bias  
 $C$  : Parameter regulasi  
 $y_i$  : Label kelas  $\{-1, +1\}$   
 $X_i$  : Vektor fitur hasil TF-IDF

b) Klasifikasi dengan *Naïve Bayes*

*Naïve Bayes* merupakan algoritma probabilistik sederhana namun efektif dalam klasifikasi teks. Proses klasifikasi sentimen pada penelitian ini dilakukan dengan menghitung seluruh probabilitas dengan berdasarkan kemunculan kata[24]. Rumus dari Teorema Bayes dapat dihitung menggunakan persanamaan.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \dots\dots\dots (2)$$

Keterangan:

- $B$  : Data sampel dengan label yang akan diklasifikasi  
 $A$  : Kelas sentiment (positif atau negatif)  
 $P(A|B)$  : Probabilitas dokumen B termasuk kelas A  
 $P(A)$  : Probabilitas awal kelas A  
 $P(B|A)$  : Probabilitas kata-kata dalam dokumen B muncul dalam kelas A  
 $P(B)$  : Probabilitas dokumen B muncul secara keseluruhan

6) **Evaluasi Model**

Setelah klasifikasi akhir, tahap selanjutnya adalah mengevaluasi kinerja model SVM dan *Naïve Bayes* yang dinilai menggunakan *confusion matrix* dan diterapkan pada data uji (20% dari dataset) untuk melihat sejauh mana model mampu mengklasifikasikan data baru berdasarkan empat komponen utama yaitu *True Positive (TP)*, *False Positive (FP)*, *True sentiment (TN)*, dan *False Negative (FN)*.

|            |          | ACTUAL            |                   |
|------------|----------|-------------------|-------------------|
|            |          | Negative          | Positive          |
| PREDICTION | Negative | TRUE<br>NEGATIVE  | FALSE<br>NEGATIVE |
|            | Positive | FALSE<br>POSITIVE | TRUE<br>POSITIVE  |

Gambar 2. *Confusion matrix*

Dari *confusion matrix* tersebut dihitung metrik evaluasi yang meliputi,

- a) *Accuracy*: untuk mengukur proporsi prediksi yang benar terhadap seluruh data uji.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots\dots\dots (3)$$

- b) *Precision (weighted average)*: untuk mengukur ketepatan prediksi model dalam memetakan suatu sampel kelas tertentu dengan mempertimbangkan masing-masing kelas.

$$Precision_{weighted} = \sum_{i=1}^k w_i \cdot \frac{TP_i}{TP_i + FP_i} \dots\dots\dots (4)$$

Keterangan:  $w$  adalah bobot kelas

- c) *Recall (weighted average)*: untuk mengukur seberapa baik model dalam menemukan seluruh sampel yang seharusnya masuk pada kelas tertentu dengan pembobotan sesuai distribusi kelas.

$$Recall_{weighted} = \sum_{i=1}^k w_i \cdot \frac{TP_i}{TP_i + FN_i} \dots\dots\dots (5)$$

- d) *F-1 Score*: rata-rata antara *precision* dan *recall* pada masing-masing kelas yang dihitung secara proporsional terhadap support tiap kelas.

$$F1_{weighted} = \sum_{i=1}^k w_i \cdot \left( 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i} \right) \dots\dots\dots (6)$$

Penggunaan *weighted average* pada penelitian ini untuk memberikan gambaran performa yang lebih representatif terhadap seluruh kelas. Dengan demikian, kombinasi antara evaluasi per kelas pada proses klasifikasi dan evaluasi keseluruhan dengan *weighted matrices* dapat memberikan pemahaman yang menyeluruh terhadap performa model.

## 4. Hasil dan Pembahasan

### 4.1 Crawling Data

Data yang dikumpulkan dari X mencakup 1159 tweet dengan menggunakan beberapa kata kunci seperti “Danantara”, “#Danantara”, dan istilah lainnya yang terkait dengan topik Danantara.

| conversation_id_str | created_at                                   | favorite_count | full_text  | id_str      | image_url                                      | in_reply_to_screen_name | lang | location                           | quote_count | reply_count | retweet_count | url                                      |
|---------------------|--|----------------|--|-------------|--|-------------------------|------|------------------------------------|-------------|-------------|---------------|--|
| 0                   | 1911791e+18<br>Mon Apr 14 15:10:10+0000 2025 | 0              | @jutasadbo3<br>@memankuy<br>@sua423<br>Woy yg p...           | 1911804e+18 | NaN  | jutasadbo3              | in   | NaN                                | 0           | 1           | 0             | https://x.com/MasQamar/status/19118041   |
| 1                   | 1911801e+18<br>Mon Apr 14 15:17:17+0000 2025 | 0              | Investasi di<br>Danantara adalah<br>investasi untuk ...      | 1911801e+18 | https://pbs.twimg.com/media/GogUz2OWAAw_Vj.jpg | NaN                     | in   | NaN                                | 0           | 0           | 0             | https://x.com/AndraRacer/status/19118001 |
| 2                   | 1911392a+18<br>Mon Apr 14 10:27:54+0000 2025 | 0              | @ch_chotimah2<br>Ada tuh<br>DANANTARA kan<br>dibuat ...      | 1911798a+18 | NaN  | ch_chotimah2            | in   | 0000000000                         | 0           | 0           | 0             | https://x.com/aziz_lio/status/19117981   |
| 3                   | 1911794e+18<br>Mon Apr 14 14:49:38+0000 2025 | 1              | Investasi Qatar<br>Untuk Danantara<br>Indonesia #Dan...      | 1911794e+18 | https://pbs.twimg.com/media/GogGDaVAAQ02.jpg   | NaN                     | in   | Dempasar<br>Selatan,<br>Indonesia  | 0           | 0           | 0             | https://x.com/facu_maxson/status/1911    |
| 4                   | 1911700a+18<br>Mon Apr 14 14:32:42+0000 2025 | 0              | Presiden Prabowo<br>Umumkan<br>Investasi USD 2<br>Mila...    | 1911700a+18 | NaN  | NaN                     | in   | Jakarta<br>Selatan,<br>DKI Jakarta | 0           | 0           | 0             | https://x.com/jintasaandoto/status/1911  |
| ...                 | ...  | ...            | ...  | ...         | ...  | ...                     | ...  | ...                                | ...         | ...         | ...           | ...                                      |
| 1160                | 1904111e+18<br>Tue Mar 25 04:03:18+0000 2025 | 0              | @democracymedia<br>Danantara Ru<br>Kondang siwa ke...        | 1904384e+18 | NaN  | democracymedia          | in   | NaN                                | 0           | 0           | 0             | https://x.com/krokro/status/1904381      |
| 1161                | 1904100a+18<br>Tue Mar 25 04:03:24+0000 2025 | 2              | @DarmasfendiD<br>@basdu_jita<br>@ch_chotimah2<br>@tiperse... | 1904383a+18 | NaN  | DarmasfendiD            | in   | wherever<br>you are                | 0           | 0           | 1             | https://x.com/grok/status/1904382380     |

Gambar 3. Hasil crawling

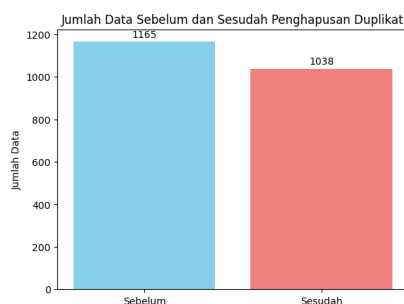
### 4.2 Preprocessing Data

*Preprocessing teks* merupakan proses pembersihan data sebelum digunakan untuk analisis. Pada tahap ini ada 5 tahapan yang termasuk *duplicate handling*. Setelah dilakukan *duplicate handling* tersisa 1038 tweet yang dapat diproses lebih lanjut.

Tabel 1. Hasil Preprocessing Data

| Proses        | Sebelum   | Sesudah   |
|---------------|---|---|
| Cleaning      | Miris dg kelakuan prabowo Katanya mau mengejar koruptor sampai keantartika. Lo kok malah ngumpulkan koruptor untuk mengelola danantara. Dari orang terkorup didunia no 2 versi OCCRP sampai buronan korupsi 15 tahun . Emaang lu wo dasar #PembohongGendut.   | miris dg kelakuan prabowo katanya mau mengejar koruptor sampai keantartika lo kok malah ngumpulkan koruptor untuk mengelola danantara dari orang terkorup didunia no versi occrp sampai buronan korupsi tahun emaang lu wo dasar pembohonggendut  |
| Tokenization  | miris dg kelakuan prabowo katanya mau mengejar koruptor sampai keantartika lo kok malah ngumpulkan koruptor untuk mengelola danantara dari orang terkorup didunia no versi occrp sampai buronan korupsi tahun emaang lu wo dasar pembohonggendut  | "miris", "dg", "kelakuan", "prabowo", "katanya", "mau", "mengejar", "koruptor", "sampai", "keantartika", "lo", "kok", "malah", "ngumpulkan", "koruptor", "untuk", "mengelola", "danantara", "dari", "orang", "terkorup", "didunia", "no", "versi", "occrp", "sampai", "buronan", "korupsi", "tahun", "emaang", "lu", "wo", "dasar", "pembohonggendut" |
| Normalization | "miris", "dg", "kelakuan", "prabowo", "katanya", "mau", "mengejar", "koruptor", "sampai", "keantartika", "lo", "kok", "malah", "ngumpulkan", "koruptor", "untuk", "mengelola", "danantara", "dari", "orang", "terkorup", "didunia", "no", "versi", "occrp", "sampai", "buronan", "korupsi", "tahun", "emaang", "lu", "wo", "dasar", | "miris", "dengan", "kelakuan", "prabowo", "katanya", "mau", "mengejar", "koruptor", "sampai", "keantartika", "lo", "kok", "malah", "ngumpulkan", "koruptor", "untuk", "mengelola", "danantara", "dari", "orang", "terkorup", "didunia", "no", "versi", "occrp", "sampai", "buronan", "korupsi", "tahun", "emaang", "lu",                              |

| Proses            | Sebelum   | Sesudah   |
|-------------------|---|---|
|                   | "pembobonggendut"   | "wo", "dasar", "pembobonggendut"  |
| Stopwords removal | "miris", "dengan", "kelakuan", "prabowo", "katanya", "mau", "mengejar", "koruptor", "sampai", "keantartika", "lo", "kok", "malah", "ngumpulkan", "koruptor", "untuk", "mengelola", "danantara", "dari", "orang", "terkorup", "didunia", "no", "versi", "occrp", "sampai", "buronan", "korupsi", "tahun", "emaang", "lu", "wo", "dasar", "pembobonggendut" | "miris", "kelakuan", "prabowo", "mengejar", "koruptor", "keantartika", "lo", "ngumpulkan", "koruptor", "mengelola", "danantara", "orang", "terkorup", "didunia", "no", "versi", "occrp", "buronan", "korupsi", "emaang", "lu", "wo", "dasar", "pembobonggendut" |
| Stemming          | "miris", "kelakuan", "prabowo", "mengejar", "koruptor", "keantartika", "lo", "ngumpulkan", "koruptor", "mengelola", "danantara", "orang", "terkorup", "didunia", "no", "versi", "occrp", "buronan", "korupsi", "emaang", "lu", "wo", "dasar", "pembobonggendut"   | "miris", "laku", "prabowo", "kejar", "koruptor", "antartika", "lo", "ngumpulkan", "koruptor", "kelola", "danantara", "orang", "korup", "dunia", "no", "versi", "occrp", "buron", "korupsi", "emaang", "lu", "wo", "dasar", "pembobonggendut"                    |

Gambar 4. Hasil penghapusan *tweet* duplikat

### 4.3 Pelabelan

Proses *labelling* dilakukan dengan menggabungkan kamus *Lexicon* dan SentiStrength\_id dengan total 3913 kosa kata positif dan 7263 kosakata negatif. Tahapan ini dilakukan sebelum *stemming* karena beberapa kosa kata yang ada di dalam kedua kamus memuat bentuk berimbuhan yang dapat hilang jika dilakukan *stemming* terlebih dahulu.

Kedua kamus memiliki rentang bobot -5 sampai dengan 5 yang kemudian di standarisasi menjadi Positif (+1) dan negatif (-1). *Tweet* dengan proporsi netral cukup kecil dan berpotensi mengganggu pembelajaran model biner, oleh karena itu untuk kategori dengan skor 0 (netral) akan di-*filter* dalam proses pelatihan dan pengujian. Data yang telah di-*filter* tersisa 770 *tweets* yang terdiri dari 466 sentimen negatif dan 304 sentimen positif.

|      | no_stopwords                                      | label_sentimen_numeric | label_sentimen_text |
|------|---|------------------------|---------------------|
| 0    | [woy, dibawa, danantara, potensi, keuntungan, ... | 1                      | Positive            |
| 1    | [investasi, danantara, investasi, kemajuan, ba... | 1                      | Positive            |
| 2    | [tuh, danantara, dibentuk, invest, programnya,... | -1                     | Negative            |
| 3    | [investasi, qatar, danantara, indonesia, danan... | 0                      | Neutral             |
| 4    | [presiden, prabowo, umumkan, investasi, usd, m... | 1                      | Positive            |
| ...  | ...   | ...                    | ...                 |
| 1160 | [danantara, kandang, tikus, tep, korupsi, hala... | -1                     | Negative            |
| 1161 | [pereode, terima, kasih, bro, bantu, jelasin, ... | 1                      | Positive            |
| 1162 | [danantara, puncak, klasemen, liga, korupsi, i... | -1                     | Negative            |
| 1163 | [miris, kelakuan, prabowo, mengejar, koruptor,... | -1                     | Negative            |
| 1164 | [namanya, dewan, penasihat, pejabat, danantara... | -1                     | Negative            |

1038 rows x 3 columns

Gambar 5. Hasil *labelling*



```
df_binner = df_labeled[df_labeled['label_sentimen_text'] != 'Neutral'].copy()

print(f"Jumlah data setelah filtering: {len(df_binner)}")

Jumlah data setelah filtering: 770
```

**Gambar 6.** Hasil *filtering*

#### 4.4 Hasil Pembobotan TF-IDF

Pada bagian ini, data sentimen yang telah melewati fase *filtering* akan dibagi menjadi data latih (80%) dan data uji (20%) dengan memastikan proporsi setiap kelas sama dengan proporsi di dataset asli. Proses pembobotan teks menggunakan TF-IDF *n-gram* (1-2) untuk mengenai kata penting (unigram) dan memahami konteks *tweet* (bigram).

| Term            | D1    | D2  | D3  | D4  | D5  | D6  | D7  | D8  | D9  | ... | D761 | D762 | D763 | D764 | D765 | D766 | D767 | D768 | D769 | D770 |
|-----------------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|
| 159 astia       | 0.241 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| 538 deviden     | 0.241 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| 622 efisiensi   | 0.241 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| 786 goal        | 0.241 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| 1780 pendidikan | 0.241 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| 2068 sebagain   | 0.241 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| 2564 woy        | 0.241 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| 32 sir          | 0.227 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| 440 cita        | 0.227 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| 298 bidang      | 0.217 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |

**Gambar 7.** Hasil TF-IDF berdasarkan beberapa dokumen pertama

Setelah pembobotan dilakukan, ditemukan adanya ketidakseimbangan data, di mana jumlah tweet sentimen negatif lebih banyak dibandingkan tweet positif. Untuk mengatasi masalah ini diterapkan Teknik SMOTE (*Syntetic Minority Oversampling Technique*) pada data latih. SMOTE bekerja dengan menghasilkan sampel sintesis untuk kelas minoritas agar kelas positif seimbang dengan negatif sebelum memasuki proses pelatihan model.

**Tabel 2.** Hasil Jumlah data SMOTE untuk *Naïve Bayes*

| Sentimen | Sebelum SMOTE | Sesudah SMOTE |
|----------|---------------|---------------|
| Negatif  | 373           | 373           |
| Positif  | 243           | 373           |

#### 4.5 Klasifikasi Model

Sebelum melakukan klasifikasi akhir, model dievaluasi menggunakan *5-fold cross validation* dengan empat scenario yaitu, SVM *baseline* (tanpa *balancing*), SVM dengan SMOTE, *Naïve Bayes baseline*, dan *Naïve Bayes SMOTE*.

**Tabel 3.** Hasil evaluasi dengan 5-fold cross validation

| Model                  | Mean F1-Weighted | Std Dev |
|------------------------|------------------|---------|
| SVM (SMOTE)            | 0.8359           | 0.0558  |
| SVM (baseline)         | 0.8444           | 0.0488  |
| Naïve Bayes (SMOTE)    | 0.8411           | 0.0283  |
| Naïve Bayes (baseline) | 0.8091           | 0.0443  |

Meskipun SVM *baseline* menunjukkan nilai *F1-Weighted* sedikit lebih tinggi dibandingkan dengan SVM (SMOTE), perbedaan tersebut tidak signifikan. Oleh karena itu, scenario dengan penerapan SMOTE tetap digunakan pada tahap evaluasi data uji untuk memastikan distribusi kelas yang lebih seimbang serta untuk menganalisis kemampuan model dalam mengenali kelas minoritas secara lebih menyeluruh.



**Tabel 4.** Hasil klasifikasi model SVM (SMOTE)

| Kelas   | Precision | Recall | F1-Score | Support |
|---------|-----------|--------|----------|---------|
| Negatif | 86,1%     | 93,5%  | 89,7%    | 93      |
| Positif | 88,7%     | 77%    | 82,5%    | 61      |

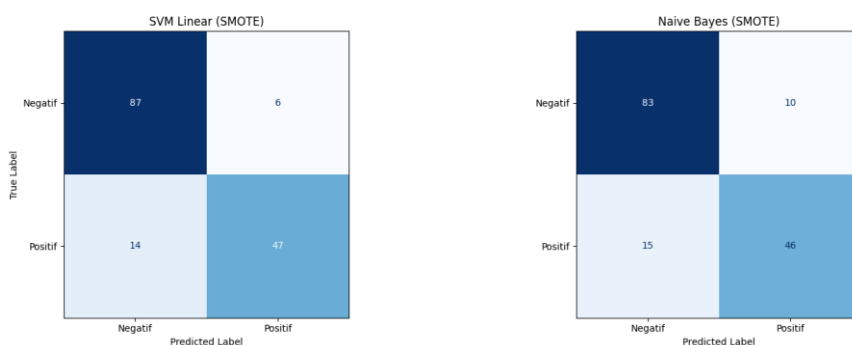
**Tabel 5.** Hasil klasifikasi model *Naïve Bayes* (SMOTE)

| Kelas   | Precision | Recall | F1-Score | Support |
|---------|-----------|--------|----------|---------|
| Negatif | 84,7%     | 89,2%  | 86,9%    | 93      |
| Positif | 82%       | 75,4%  | 78,6%    | 61      |

Berdasarkan hasil klasifikasi pada tabel 4 dan 5, model *Support Vector Machine* (SVM) dengan kernel linear dengan penerapan SMOTE menunjukkan performa klasifikasi yang lebih unggul dibandingkan *Naïve Bayes*. SVM memperoleh akurasi sebesar 87% dengan nilai precision, recall, dan f1-score yang relatif seimbang. Namun, baik model SVM dan *Naïve Bayes* menunjukkan pola yang sama dalam menghadapi kesulitan mengenali kelas positif, hal ini terlihat pada nilai *recall* pada kedua model yang lebih rendah dibandingkan kelas negatif. Kondisi ini menunjukkan bahwa adanya penerapan SMOTE mampu menyeimbangkan distribusi kelas, namun belum sepenuhnya meningkatkan kemampuan model dalam mengenali kelas positif dengan baik yang dipengaruhi oleh karakteristik kata pada kelas positif yang lebih beragam.

#### 4.6 Evaluasi Model

Confusion Matrices for SVM SMOTE and Naive Bayes SMOTE

**Gambar 8.** Confusion matrix SVM (Kiri) dan *Naïve Bayes* (Kanan)

Berdasarkan *confussion matrix* di atas akan dihitung *accuracy precision*, *recall*, dan *f1-score* dengan merujuk pada Tabel 4 dan 5:

a) Accuracy:

1) *Support Vector Machine* (SVM)

$$\text{Accuracy}_{\text{SVM}} = \frac{\text{TP} + \text{TN}}{N_{\text{Total}}} = \frac{47 + 87}{154} = \frac{134}{154} \approx 0.8701$$

2) *Naïve Bayes*

$$\text{Accuracy}_{\text{NB}} = \frac{\text{TP} + \text{TN}}{N_{\text{Total}}} = \frac{46 + 83}{154} = \frac{129}{154} \approx 0.8377$$

b) Precision:

1) *Support Vector Machine* (SVM)

$$(0,861 \times 93 + 0,887 \times 61) / 154 = 0.871 \quad (9)$$

2) *Naïve Bayes*

$$(0,847 \times 93 + 0,821 \times 61) / 154 = 0.837 \quad (10)$$

c) Recall:

1) *Support Vector Machine* (SVM)

$$(0,935 \times 93 + 0,770 \times 61) / 154 = 0.870 \quad (11)$$

- 2) Naïve Bayes  
 $(0,892 \times 93 + 0,754 \times 61) / 154 = 0.838$  (12)
- d) F1-Score:
- 1) *Support Vector Machine* (SVM)  
 $(0,897 \times 93 + 0,825 \times 61) / 154 = 0.868$  (13)
- 2) *Naïve Bayes*  
 $(0,869 \times 93 + 0,786 \times 61) / 154 = 0.836$  (14)

**Tabel 6.** Hasil Evaluasi

| Model                         | Accuracy | Precision | Recall | F1-Score |
|-------------------------------|----------|-----------|--------|----------|
| <i>Support Vector Machine</i> | 87%      | 87,1%     | 87%    | 86,8%    |
| <i>Naïve Bayes</i>            | 83,7%    | 83,7%     | 83,8%  | 83,6%    |

Berdasarkan hasil evaluasi pada Tabel 6, terlihat bahwa model *Support Vector Machine* (SVM) dengan kernel linear memiliki performa lebih unggul dibandingkan dengan *Naïve Bayes*. SVM memperoleh akurasi sebesar 87%, lebih tinggi dibandingkan *Naïve Bayes* yang menghasilkan 83,7%. Perbedaan ini juga tercermin pada metrik lain seperti *precision*, SVM memiliki nilai 87,1% yang artinya memiliki kemampuan yang lebih baik dalam mengidentifikasi tweet sebagai sentimen positif. Sebaliknya, *Naïve Bayes* memiliki nilai *precision* sebesar 83,7% yang artinya model ini lebih sering membuat *false positive*.

Dari sisi *recall*, terlihat bahwa SVM memiliki nilai *recall* 87% sedangkan *Naïve Bayes* 83,8%. Nilai *recall* tersebut menunjukkan seberapa baik kedua model mengenali seluruh tweet yang benar-benar termasuk ke dalam kelas positif. Dalam konteks ini, SVM lebih sensitif dalam mendeteksi tweet *True Positive* dan menghasilkan *False Negative* yang lebih sedikit dibandingkan dengan *Naïve Bayes*. Perbedaan nilai *precision* dan *recall* mempengaruhi nilai F1-Score yang menandakan bahwa SVM memiliki keseimbangan yang lebih baik antara *precision* dan *recall*.

Secara keseluruhan dari hasil evaluasi pada studi kasus Danantara menunjukkan bahwa SVM merupakan algoritma yang lebih stabil dan akurat dalam melakukan klasifikasi sentimen dibandingkan dengan *Naïve Bayes*.

#### 4.7 Pembahasan

Berdasarkan hasil penelitian, algoritma *Support Vector Machine* (SVM) memberikan performa klasifikasi sentimen yang lebih unggul dibandingkan dengan *Naïve Bayes* pada data tweet terkait program Danantara. Model SVM menghasilkan nilai akurasi sebesar 87%, lebih tinggi dibandingkan *Naïve Bayes* sebesar 83,7% serta menunjukkan keseimbangan *precision* dan *recall* yang lebih baik pada kedua kelas sentiment.

Temuan ini memperkuat hasil penelitian Nugraha[2] yang menyatakan bahwa sentimen masyarakat terhadap Danantara cenderung didominasi oleh sentimen negatif. Adanya kecenderungan persebaran sentimen negatif disebabkan karena kosakata negatif pada kamus *lexicon* lebih banyak dibandingkan kosakata positif. Berbeda dengan penelitian Nugraha yang murni menggunakan *lexicon-based*, penelitian ini menunjukkan bahwa integrasi pelabelan *lexicon* dengan algoritma *machine learning* mampu menghasilkan evaluasi performa yang lebih terukur.

Perbedaan capaian performa antara penelitian ini dan penelitian Hermanto et.al.[19] tidak hanya dipengaruhi oleh algoritma klasifikasi yang digunakan, tetapi juga teknik pelabelan data. Hermanto et al. menerapkan pelabelan otomatis berbasis Pustaka Deberta, yaitu model *pre-trained deep learning* yang menentukan label sentimen berdasarkan representasi kontekstual kalimat hasil pembelajaran internal. Meskipun pendekatan tersebut efektif dalam menangkap konteks linguistik secara menyeluruh, mekanisme penentuan label tidak secara eksplisit menunjukkan bobot kata tertentu. Sebaliknya, penelitian ini menggunakan pelabelan *hybrid* berbasis InSet dan SentiStrength\_id yang dirancang khusus untuk Bahasa Indonesia, sehingga proses pelabelan dapat dijelaskan melalui kosakata yang digunakan. Pendekatan ini memungkinkan interpretasi linguistik yang lebih jelas terhadap faktor-faktor yang memengaruhi hasil klasifikasi dan berkontribusi pada peningkatan performa model SVM yang diperoleh dalam penelitian ini.

Selain itu, hasil penelitian ini sejalan dengan temuan Kurnia [20] yang menyatakan bahwa penerapan SMOTE mampu meningkatkan performa *Naïve Bayes* pada data yang

mengalami ketidakseimbangan kelas. Penelitian ini menunjukkan adanya peningkatan performa pada algoritma *Naïve Bayes* setelah diterapkan SMOTE meskipun hasil akurasi masih di bawah SVM. Hal ini menunjukkan meskipun SMOTE efektif untuk menstabilkan pembelajaran *Naïve Bayes*, algoritma tersebut masih memiliki keterbatasan dalam menangkap konteks antar kata pada data teks pendek seperti tweet.

Kontribusi penelitian ini terletak pada penyajian perbandingan yang komprehensif antara algoritma antara algoritma *Support Vector Machine* dan *Naïve Bayes* dalam analisis sentimen terhadap Danantara dengan menggunakan pelabelan *hybrid* berbasis InSet Lexicon dan SentiStrength\_id. Penelitian ini menunjukkan bahwa pendekatan pelabelan *lexicon* dapat digunakan untuk klasifikasi awal data latih bagi model *supervised learning* serta menunjukkan adanya perbedaan respon algoritma terhadap ketidakseimbangan kelas yang mana *Naïve Bayes* memperoleh manfaat yang besar dari penerapan SMOTE, sementara SVM mempertahankan performa yang konsisten pada skenario pelatihan yang digunakan.

Meskipun pendekatan *hybrid lexicon* yang digunakan dapat menghasilkan performa klasifikasi yang terukur, penelitian ini memiliki keterbatasan pada proses pelabelan yang sepenuhnya bersifat otomatis dan bergantung pada distribusi kosakata dalam InSet Lexicon dan SentiStrength\_id. Dominasi kosakata negatif berpotensi memengaruhi sentiment serta menurunkan kemampuan model dalam mengenali kelas positif. Oleh karena itu, penelitian selanjutnya disarankan untuk melakukan pelabelan secara manual agar interpretasi sentimen dapat lebih representatif.

## 5. Simpulan

Penelitian ini memiliki tujuan untuk menganalisis sentimen masyarakat Indonesia di platform X terhadap program Danantara serta membandingkan performa algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* dalam mengklasifikasikan sentimen berdasarkan data *tweet*. Berdasarkan keseluruhan rangkaian proses mulai dari *crawling* hingga evaluasi diperoleh beberapa simpulan penting. Pendekatan *hybrid* yang menggabungkan InSet Lexicon dan SentiStrength\_id berhasil memberikan dasar pelabelan yang kuat untuk menangani ragam bahasa informal. Namun, distribusi kosakata pada kedua kamus cenderung memuat kosakata negatif daripada positif yang dapat mengenali konteks *tweet* negatif dalam dataset lebih sering dan menimbulkan ketidakseimbangan kelas. Oleh karena itu, SMOTE menjadi penting untuk meningkatkan stabilitas pelatihan model.

Berdasarkan hasil evaluasi akhir, perbandingan kinerja kedua algoritma menunjukkan bahwa SVM unggul dalam mengklasifikasikan sentimen dibandingkan dengan *Naïve Bayes*. Hasil pengujian menunjukkan SVM memperoleh akurasi sebesar 87%, sedangkan *Naïve Bayes* mencapai 83,7%. SVM juga menunjukkan nilai *precision*, *recall*, dan *f1-score* yang lebih seimbang karena kemampuannya memanfaatkan fitur TF-IDF yang berdimensi tinggi dan menemukan *hyperplane* optimal pada data yang memiliki variasi linguistik tinggi. Di lain sisi, *Naïve Bayes* lebih terbatas karena asumsi independensi fitur dan kesulitannya menangkap konteks bigram sehingga lebih rentan terhadap kesalahan klasifikasi terutama ketika menghadapi ekspresi positif. Peningkatan kinerja *Naïve Bayes* setelah SMOTE menunjukkan bahwa algoritma ini tetap relevan sebagai baseline model klasifikasi sentimen.

Secara keseluruhan, penelitian ini menyimpulkan bahwa algoritma SVM lebih unggul untuk analisis sentimen berbasis media sosial terhadap isu kebijakan publik seperti Danantara. Pendekatan *hybrid Lexicon-based* ditambah SMOTE terbukti dapat memperbaiki kualitas label dan meningkatkan performa klasifikasi. Temuan penting lainnya yaitu dominasi kosakata negatif dalam kamus InSet + SentiStrength\_id turut memengaruhi distribusi awal sentimen, sehingga penelitian selanjutnya disarankan untuk memperluas kosakata positif atau melakukan pelabelan secara manual agar interpretasi sentimen dapat lebih representatif.

## Daftar Referensi

- [1] Kementerian Keuangan Indonesia, "Presiden Prabowo Resmi Luncurkan Badan Pengelola Investasi Danantara," 2025. [Online]. Tersedia: <https://www.kemenkeu.go.id/informasi-publik/publikasi/berita-utama/Presiden-Luncurkan-Danantara> [Diakses:15 Juni 2025]
- [2] S. A. 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: <https://doi.org/10.36040/jati.v9i3.13836>.
- [3] Pemerintah Republik Indonesia, "Peraturan Pemerintah Republik Indonesia Nomor 10

- Tahun 2025 Tentang Organisasi dan Tata Kelola Badan Pengelola Investasi Daya Anagata Nusantara.* 2025.
- [4] M. P. Pulungan, A. Purnomo, A. Kurniasih, P. Korespondensi, I. Class, and S. M. O. Technique, "Penerapan Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Kepribadian MbtI Menggunakan Naive Bayes," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 11, no. 5, pp. 1033–1042, 2024, doi: 10.25126/jtiik.2024117989.
  - [5] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA J. STIKI Inform. J.*, vol. 10, no. 2, pp. 71–76, 2020.
  - [6] D. D. Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat ( DPR ) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: <https://doi.org/10.23960/jitet.v10i1.2262>.
  - [7] R. Ramlan, N. Satyahadewi, and W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM," *Jambura J. Math.*, vol. 5, no. 2, pp. 431–445, 2023, doi: 10.34312/jjom.v5i2.20860.
  - [8] A. Corallo *et al.*, "Sentiment analysis for government: An optimized approach," in *Proceedings of the International Conference on Electronic Government and the Information Systems Perspective, Lecture Notes in Computer Science*, vol. 9166, no. 7, Springer, pp. 98–112, 2015, doi: 10.1007/978-3-319-21024-7\_7.
  - [9] N. Hadi and D. Sugiarto, "Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM , Logistic Regression dan Naive Bayes," *J. Inform. J. Pengemb. IT*, vol. 10, no. 1, pp. 37–49, 2025, doi: 10.30591/jpit.v10i1.7106.
  - [10] W. Ningsih, B. Alfianda, R. Rahmadden, and D. Wulandari, "Perbandingan Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.
  - [11] 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.
  - [12] D. W. Syahputra, B. Rahayudi, and L. Muflikhah, "Analisis Sentimen Twitter terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat menggunakan Metode Support Vector Machine," *Fak. Ilmu Komputer, Univ. Brawijaya*, vol. 6, no. 3, pp. 1067–1072, 2022.
  - [13] T. Widyanto, I. Ristiana, and A. Wibowo, "Komparasi Naive Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter," *SINTECH J.*, vol. 6, no. 3, pp. 147–161, 2023, doi: <https://doi.org/10.31598/sintechjournal.v6i3.1433>.
  - [14] A. P. Astuti, S. Alam, and I. Jaelani, "Komparasi Algoritma Support Vector Machine Dengan Naive Bayes Untuk Analisis Sentimen Pada Aplikasi BRImo," *Jurnal Bangkit Indonesia*, vol. XI, no. 02, pp. 1–7, 2023.
  - [15] A. R. Isnain *et al.*, "Comparison of Support Vector Machine and Naive Bayes on Twitter Data Sentiment Analysis," *J. Inform. J. Pengemb. IT*, vol. 6, no. 1, pp. 56–60, 2021.
  - [16] F. A. Ryandi, D. Pratiwi, and S. Sari, "Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes dan SVM," *J. Sains dan Teknol.*, vol. 7, no. 1, pp. 1–6, 2025.
  - [17] F. Koto, "InSet Lexicon : Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs InSet Lexicon : Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs," in *Proc. Int. Conf. on Asian Language Processing (IALP)*, pp. 391–396, 2017, doi: 10.1109/IALP.2017.8300625..
  - [18] D. H. Wahid, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 10, no. 2, pp. 207–218, 2017.
  - [19] M. L. Hermanto, Fathoni, O. Ardhillah, and A. Ibrahim, "Analisis Sentimen Masyarakat Terhadap Danantara Di Platform X Dengan Metode SVM," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 6779–6785, 2025, doi: <https://doi.org/10.36040/jati.v9i4.14189>.
  - [20] Kurnia, I. Purnamasari, and D. D. Saputra, "Analisis Sentimen dengan Metode Naive Bayes , SMOTE dan Adaboost pada Twitter Bank BTN," *JTIK (Jurnal Teknol. Inf. dan*

- Komunikasi*), vol. 7, no. 2, pp. 236–242, 2023, doi: <https://doi.org/10.35870/jtik.v7i3.707>
- [21] D. Purnamasari *et al.*, *Pengantar Metode Analisis Sentimen*. Depok: Gunadarma Penerbit, 2023.
- [22] M. K. Sandryan, B. Rahayudi, and D. E. Ratnawati, “Analisis Sentimen Pada Media Sosial Twitter Terhadap Undang-Undang Cipta Kerja Menggunakan Algoritma Backpropagation dan Term Frequency-Inverse Document Frequency,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5349–5355, 2021.
- [23] J. W. Iskandar and Y. Nataliani, “Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek Jessica,” *J. RESTI*, vol. 5, no. 6, pp. 1120–1126, 2021.
- [24] M. Safrudin, M. Martanto, and U. Hayati, “Perbandingan Kinerja Naïve Bayes Dan Support Vector Machine Untuk Klasifikasi Sentimen Ulasan Game Genshin Impact,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3182–3188, 2024, doi: 10.36040/jati.v8i3.8415.