

Analisis Sentimen Pengguna X terhadap Kasus Korupsi Gula Tom Lembong Menggunakan Naïve Bayes, SVM, dan Random Forest

Aneira Vicentiya Kuncoro*, Fikri Budiman, Defri Kurniawan

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ^{1,*}111202214226@mhs.dinus.ac.id, ²fikri.budiman@dsn.dinus.ac.id, ³defri.kurniawan@dsn.dinus.ac.id

Email Penulis Korespondensi: 111202214226@mhs.dinus.ac.id

Submitted: 21/10/2025; Accepted: 06/12/2025; Published: 08/12/2025

Abstrak—Kasus dugaan korupsi impor gula yang melibatkan Tom Lembong menjadi salah satu isu yang banyak diperbincangkan publik di media sosial dan menimbulkan beragam reaksi. Fenomena ini menunjukkan bagaimana opini masyarakat terhadap isu hukum sering kali dipengaruhi oleh persepsi terhadap tokoh publik yang terlibat. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kasus tersebut di platform media sosial X (Twitter). Data penelitian terdiri dari 1.802 tweet yang dikumpulkan melalui proses crawling menggunakan API X dengan kata kunci “Tom Lembong”. Tahapan penelitian meliputi data cleaning, case folding, normalisasi teks, tokenizing, stopword removal, stemming, pelabelan sentimen menggunakan pendekatan lexicon-based, serta ekstraksi fitur dengan metode Term Frequency–Inverse Document Frequency (TF-IDF). Dataset yang telah siap kemudian diuji menggunakan tiga algoritma klasifikasi, yaitu Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Hasil pengujian menunjukkan bahwa algoritma SVM memiliki performa terbaik dengan akurasi 84%, diikuti oleh Random Forest dengan 80% dan Naïve Bayes dengan 76%. Berdasarkan hasil pelabelan, sentimen positif mendominasi dengan persentase 61%, sedangkan sentimen negatif sebesar 39%. Meskipun isu yang dianalisis berkaitan dengan dugaan korupsi, dominasi sentimen positif mengindikasikan bahwa opini publik cenderung berfokus pada citra personal atau rekam jejak publik Tom Lembong yang dinilai positif, bukan semata pada substansi tuduhan hukum yang berkembang. Temuan ini menunjukkan efektivitas algoritma SVM dalam menganalisis teks berdimensi tinggi serta memberikan wawasan tentang bagaimana persepsi publik terhadap isu hukum dapat dipengaruhi oleh faktor citra dan konteks sosial politik di media sosial.

Kata Kunci: Analisis Sentimen; Media Sosial X; Tom Lembong; Support Vector Machine; Random Forest; Naïve Bayes

Abstract—The alleged sugar import corruption case involving Tom Lembong has become one of the most widely discussed public issues on social media, generating diverse reactions. This phenomenon illustrates how public opinion on legal issues is often influenced by perceptions of the public figures involved. This study aims to analyze public sentiment regarding the case on the social media platform X (formerly Twitter). The dataset consists of 1,802 tweets collected through a crawling process using the X API with the keyword “Tom Lembong.” The research stages include data cleaning, case folding, text normalization, tokenizing, stopword removal, stemming, sentiment labeling using a lexicon-based approach, and feature extraction with the Term Frequency–Inverse Document Frequency (TF-IDF) method. The prepared dataset was then tested using three classification algorithms: Naïve Bayes, Support Vector Machine (SVM), and Random Forest. The results show that the SVM algorithm achieved the highest accuracy (84%), followed by Random Forest (80%) and Naïve Bayes (76%). Based on the sentiment labeling results, positive sentiment dominated with 61%, while negative sentiment accounted for 39%. Although the analyzed issue concerns an alleged corruption case, the dominance of positive sentiment indicates that public opinion tends to focus on Tom Lembong’s personal image or public track record, which is viewed positively rather than on the substance of the legal allegations. These findings demonstrate the effectiveness of the SVM algorithm in analyzing high-dimensional text and provide insights into how public perception of legal issues can be influenced by image factors and the socio-political context on social media.

Keywords: Sentiment Analysis; Social Media X; Tom Lembong; Support Vector Machine; Random Forest; Naïve Bayes

1. PENDAHULUAN

Korupsi merupakan salah satu permasalahan serius yang terus menghantui Indonesia karena berdampak pada stabilitas ekonomi, kepercayaan publik, serta tata kelola pemerintahan. Salah satu kasus yang menimbulkan perhatian publik adalah dugaan korupsi impor gula yang melibatkan mantan Menteri Perdagangan, Tom Lembong. Jaksa menuduh Tom Lembong menyetujui impor gula kristal mentah tanpa melalui prosedur koordinasi antar kementerian meskipun Indonesia mengalami surplus gula, dengan melibatkan sejumlah perusahaan swasta [1]. Perbuatan ini menurut dakwaan menyebabkan kerugian negara sebesar Rp 578 miliar, meskipun majelis hakim kemudian hanya mengakui kerugian sekitar Rp 194 miliar. Pada akhirnya, Tom Lembong divonis 4 tahun 6 bulan penjara dan denda Rp 750 juta pada tanggal 18 Juli 2025, meskipun hakim menyatakan bahwa ia tidak menikmati hasil korupsi. Perbedaan angka kerugian serta vonis yang lebih ringan daripada tuntutan menimbulkan kontroversi publik, terlebih lagi adanya kritik terhadap metode perhitungan kerugian negara oleh BPKP yang dianggap tidak akurat [2], [3], [4].

Perkembangan teknologi digital menjadikan X (sebelumnya Twitter) sebagai kanal utama bagi masyarakat dalam menyampaikan opini mengenai kasus-kasus aktual. Analisis sentimen adalah metode yang digunakan untuk mengidentifikasi, mengolah, dan mengklasifikasikan opini publik berdasarkan teks, sehingga dapat diketahui kecenderungan emosi atau sikap pengguna terhadap suatu isu. Dengan bantuan *machine learning*, analisis sentimen memungkinkan klasifikasi opini publik menjadi sentimen positif dan negatif sehingga dapat menggambarkan persepsi masyarakat terhadap isu tertentu. Dengan demikian, platform X menjadi sumber data yang relevan untuk menilai persepsi publik terhadap kasus hukum seperti dugaan korupsi Tom Lembong [5], [6], [7].

Berbagai penelitian terdahulu telah memanfaatkan analisis sentimen untuk isu-isu kebijakan publik maupun fenomena sosial. Anjana Zharifa dkk. menganalisis sentimen masyarakat pasca debat kelima pilpres 2024 dengan

Naïve Bayes dengan akurasi 81% [7]. Gina Purnama dkk. membandingkan algoritma *Support Vector Machine* dan *Random Forest* dalam analisis sentimen pinjaman online di media sosial, di mana *Random Forest* menunjukkan akurasi tertinggi sebesar 100% selisih 1% SVM dengan 99% [8]. Penelitian Elsa Triningsih dkk. mengungkap dominasi sentimen negatif terhadap Program Makan Bergizi Gratis, dengan SVM menghasilkan akurasi tertinggi sebesar 85,74% [9]. Sementara itu, Ervin Danuarta dkk. membandingkan algoritma *Naïve Bayes* dan *Random Forest* dalam analisis sentimen terhadap isu *cyberbullying* di kalangan Generasi Z pada Twitter dengan *Random Forest* yang unggul 94% [10].

Berdasarkan berbagai penelitian terdahulu, analisis sentimen telah banyak dimanfaatkan untuk memahami persepsi masyarakat terhadap kebijakan, isu publik, maupun fenomena sosial melalui data media sosial. Meskipun demikian, hingga saat ini belum ditemukan penelitian yang secara khusus membahas sentimen publik terhadap kasus dugaan korupsi impor gula yang melibatkan Tom Lembong. Sejumlah studi menunjukkan bahwa algoritma *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* memiliki kinerja yang baik dalam klasifikasi teks di berbagai konteks, seperti ulasan produk, opini publik, dan komentar di media sosial. Namun, ketiga algoritma tersebut belum pernah dibandingkan secara langsung menggunakan dataset yang berkaitan dengan kasus hukum di Indonesia, khususnya kasus dugaan korupsi impor gula.

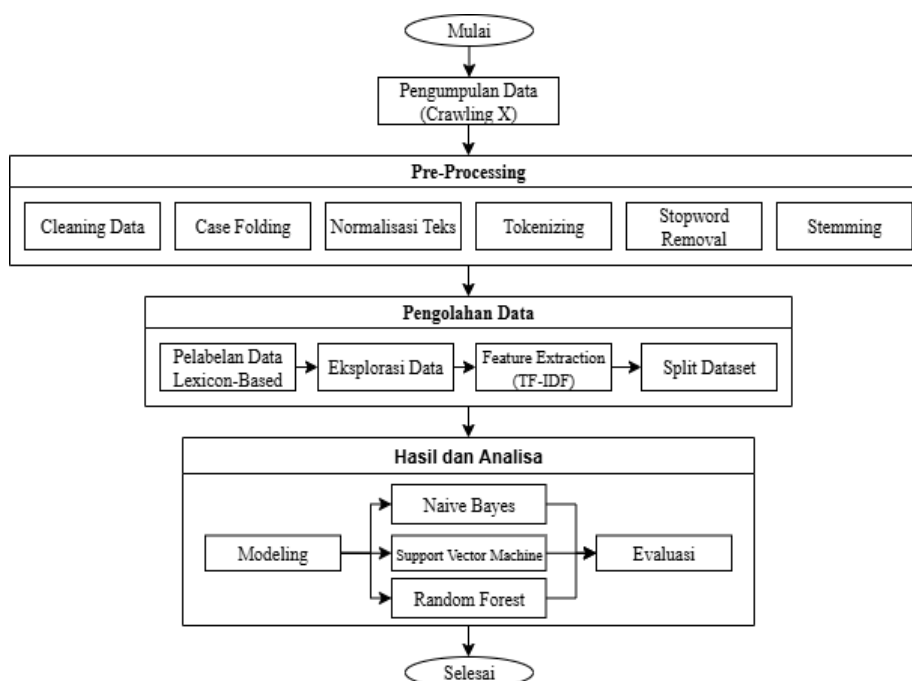
Dalam penelitian ini, pelabelan sentimen dilakukan menggunakan pendekatan *lexicon-based* (kamus sentimen), yang menghasilkan label otomatis berdasarkan daftar kata dan aturan tertentu tanpa melibatkan anotasi manual. Metode ini dipilih karena efisien dan konsisten dalam menangani data berjumlah besar, meskipun memiliki keterbatasan karena model *machine learning* cenderung mereplikasi pola kamus tanpa sepenuhnya memahami konteks teks, karena label berasal dari *lexicon-based*, model berpotensi hanya meniru pola kamus tanpa memahami konteks secara mendalam. Untuk mengatasi hal tersebut, diterapkan pra-pemrosesan teks (normalisasi, *stemming*, dan penghapusan kata tidak relevan), pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF), serta pembagian data latih dan data uji guna meningkatkan generalisasi model. Evaluasi dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk menilai keseimbangan performa setiap algoritma. Dengan demikian, meskipun menggunakan sistem pelabelan *lexicon-based*, penelitian ini tetap mampu menggambarkan efektivitas masing-masing algoritma dalam mengklasifikasikan sentimen publik.

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat di media sosial X terhadap kasus dugaan korupsi impor gula Tom Lembong dengan menggunakan tiga algoritma *machine learning*: *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest*. Hasil analisis penelitian ini diharapkan dapat memberikan gambaran yang lebih baik mengenai persepsi masyarakat terhadap kasus dugaan korupsi impor gula yang melibatkan Tom Lembong. Selain itu, penelitian ini diharapkan mampu memperkaya literatur terkait pemanfaatan analisis sentimen berbasis *machine learning* pada isu-isu hukum di Indonesia yang hingga kini masih jarang dilakukan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Alur lengkap proses penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

Gambar 1 menunjukkan diagram alur metode penelitian. Penelitian dimulai dengan pengumpulan data melalui *crawling* di platform X (Twitter), dilanjutkan *pre-processing* seperti pembersihan data, *case folding*, normalisasi teks, *tokenizing*, *stopword removal*, dan *stemming*. Data kemudian diberi label berbasis leksikon, dieksplorasi, dan fitur diekstraksi menggunakan TF-IDF sebelum dibagi menjadi data latih dan uji. Selanjutnya dilakukan pemodelan dengan algoritma *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest*, kemudian dievaluasi untuk menentukan kinerja terbaik, dan penelitian diakhiri dengan analisis hasil.

2.2 Pengumpulan Data

Penelitian ini berfokus pada reaksi masyarakat di media sosial X terhadap kasus dugaan korupsi impor gula oleh Tom Lembong. Data yang dikumpulkan berupa tweet yang memuat opini, komentar, atau informasi yang berkaitan langsung dengan topik tersebut. Pengumpulan data dilakukan menggunakan teknik *crawling* melalui API X dengan auth token sebagai autentikasi resmi. Data diambil menggunakan kata kunci “Tom Lembong” dari Januari-September 2025, kemudian disimpan dalam format CSV untuk tahap analisis selanjutnya [11], [12], [13]. Untuk kolom pengumpulan dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi Kolom Dataset

No	Nama Kolom	Tipe Data	Deskripsi
1	conversation_id_str	Integer	ID percakapan unik dari setiap tweet.
2	created_at	Object	Tanggal dan waktu tweet dipublikasikan.
3	favorite_count	Integer	Jumlah tanda suka (“like”) pada tweet.
4	full_text	Object	Teks lengkap dari tweet yang dianalisis.
5	id_str	Integer	ID unik untuk setiap tweet.
6	image_url	Object	URL gambar yang disertakan pada tweet.
7	in_reply_to_screen_name	Object	Username pengguna yang dibalas dalam tweet.
8	lang	Object	Kode bahasa yang digunakan dalam tweet (misalnya, “id”).
9	location	Float	Informasi lokasi pengguna (Jika ada).
10	quote_count	Integer	Jumlah tweet lain yang mengutip tweet tersebut.
11	reply_count	Integer	Jumlah balasan terhadap tweet.
12	retweet_count	Integer	Jumlah retweet yang diterima tweet.
13	tweet_url	Object	URL langsung ke tweet.
14	user_id_str	Integer	ID unik pengguna pembuat tweet.
15	username	Float	Nama pengguna (Jika tercantum).

Tabel 1 merupakan deskripsi rinci mengenai setiap atribut yang terdapat dalam dataset, meliputi nama kolom, tipe data, tingkat kelengkapan, dan fungsinya dalam penelitian.

2.3 Pre-Processing

Proses *pre-processing* dilakukan untuk mempersiapkan data teks agar layak digunakan pada tahap analisis sentimen. Adapun tahapan yang dilakukan meliputi :

a. Cleaning Data

Tahap ini bertujuan untuk membersihkan teks dari elemen-elemen yang tidak relevan seperti URL, tanda baca, angka, *mention*, *hashtag*, serta karakter khusus lainnya. Dengan demikian, data yang diolah hanya berisi teks yang bermakna [14].

b. Case Folding

Seluruh huruf pada teks diubah menjadi huruf kecil (*lowercase*) untuk menjaga konsistensi data dan menghindari perbedaan interpretasi akibat perbedaan kapitalisasi huruf [15].

c. Normalisasi Teks

Proses ini dilakukan dengan mengganti kata tidak baku, singkatan, atau bentuk slang menjadi kata baku sesuai kamus kata baku dan kamus singkatan yang digunakan dalam penelitian [16].

d. Tokenizing

Pada tahap ini, teks yang telah dinormalisasi dipecah menjadi potongan kata atau *token* agar dapat dianalisis secara lebih mendalam pada setiap unsur katanya [17].

e. Stopword Removal

Tahap ini menghapus kata-kata umum yang tidak memiliki makna signifikan terhadap analisis, seperti “dan”, “yang”, “di”, dan kata sejenis lainnya [18].

f. Stemming

Proses stemming dilakukan untuk mengubah setiap kata menjadi bentuk dasarnya menggunakan algoritma *stemming* Bahasa Indonesia, misalnya dengan pustaka *Sastrawi* [10].

2.4 Pengolahan Data

Tahap pengolahan data dilakukan dengan tujuan untuk memberikan label sentimen, mengekstraksi fitur, dan menyiapkan data agar dapat digunakan dalam proses pelatihan model [9], [18]. Adapun tahapan yang dilakukan meliputi :

a. Pelabelan Data

Pada tahap ini, setiap teks diberi label sentimen (positif, negatif) menggunakan pendekatan *lexicon-based*. Proses ini dilakukan dengan mencocokkan kata-kata dalam teks dengan kamus sentimen Bahasa Indonesia yang telah memiliki skor polaritas [18]. Nilai total polaritas dari kata-kata tersebut digunakan untuk menentukan label akhir pada setiap tweet [11].

b. Eksplorasi Data

Tahap ini bertujuan untuk memahami karakteristik data secara menyeluruh melalui analisis deskriptif, seperti jumlah data per kelas sentimen, panjang rata-rata teks, serta frekuensi kemunculan kata [16]. Hasil eksplorasi digunakan untuk melihat distribusi data dan mendeteksi adanya ketidakseimbangan (*imbalanced class*) dan *wordcloud* [9].

c. Feature Extraction (TF-IDF)

Setelah data diberi label, teks diubah menjadi representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Metode ini mengukur pentingnya suatu kata dalam dokumen dengan mempertimbangkan seberapa sering kata tersebut muncul di satu teks dibandingkan dengan seluruh korpus [14], [19], [20].

$$W_{ij} = tf_{ij} \times \log\left(\frac{N}{df_i}\right) \quad (1)$$

Persamaan (1) menjelaskan bahwa W_{ij} merupakan bobot suatu kata i pada dokumen j , dimana tf_{ij} menunjukkan frekuensi kemunculan kata i dalam dokumen j , sedangkan df_i menggambarkan jumlah dokumen yang mengandung kata tersebut. Adapun N merepresentasikan total keseluruhan dokumen dalam korpus.

d. Split Dataset

Tahap terakhir yaitu membagi data menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*testing set*) [14]. Pembagian ini bertujuan agar model dapat dilatih menggunakan sebagian data dan dievaluasi menggunakan data lain yang belum pernah dilihat sebelumnya, sehingga hasil prediksi menjadi lebih objektif [10].

2.5 Hasil dan Analisa

a. Modeling

Pada tahap ini, data yang telah melalui proses *feature extraction* dan split dataset digunakan untuk pelatihan model. Tiga algoritma yang digunakan dalam penelitian ini adalah *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* [19]. Masing-masing algoritma dipilih karena memiliki karakteristik yang berbeda dalam mengolah data teks dan sering digunakan dalam penelitian analisis sentimen.

b. Penerapan Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes Classifier* merupakan metode klasifikasi berbasis probabilitas yang memprediksi kejadian masa depan menggunakan data historis. Metode ini mengasumsikan tiap atribut bersifat independen, sehingga proses klasifikasi menjadi sederhana namun tetap efektif [21], [22], [23]. Rumusnya dinyatakan sebagai :

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (2)$$

Persamaan (2) dimana $P(C|X)$ merupakan probabilitas kelas C dengan kondisi fitur X , $P(C|X)$ adalah probabilitas fitur X jika diketahui kelas C , $P(C)$ adalah probabilitas awal kelas, dan $P(X)$ adalah probabilitas total dari fitur tersebut.

c. Penerapan Algoritma *Support Vector Machine* (SVM)

Algoritma *Support Vector Machine* (SVM) merupakan metode klasifikasi yang bertujuan mencari hyperplane optimal sebagai pemisah antar kelas. Pendekatan ini bekerja dengan memaksimalkan jarak (margin) antar kelas agar hasil klasifikasi lebih akurat [11], [17], [24]. Persamaan dasar SVM ditunjukkan pada persamaan (3) :

$$f(x) = w^T x + b \quad (3)$$

dengan w sebagai vektor bobot, x sebagai vektor fitur, dan b sebagai bias. Nilai tanda dari $f(x)$ menentukan kelas data, sedangkan fungsi *kernel* digunakan untuk mentransformasikan data ke ruang berdimensi lebih tinggi agar pemisahan kelas lebih optimal.

d. Penerapan Algoritma *Random Forest*

Algoritma *Random Forest* merupakan metode klasifikasi berbasis *ensemble learning* yang menggabungkan beberapa *decision tree* untuk meningkatkan akurasi prediksi. Setiap pohon keputusan dilatih pada subset data yang berbeda, dan hasil akhirnya ditentukan melalui proses *majority voting* [10], [25], [26]. Persamaan umum *Random Forest* dapat dituliskan pada persamaan (4) :



$$\hat{y} = mode(h_1(x), h_2(x), \dots, h_n(x)) \tag{4}$$

di mana $h_i(x)$ merupakan hasil prediksi dari setiap *decision tree* ke- i , dan \hat{y} adalah hasil akhir berdasarkan suara terbanyak.

e. Evaluasi Model

Evaluasi model bertujuan untuk menilai sejauh mana algoritma mampu mengklasifikasikan sentimen dengan benar. Pengukuran dilakukan menggunakan *confusion matrix* dengan metrik utama *accuracy*, *precision*, *recall*, dan *f1-score* [16], [25], [27].

Tabel 2. Model Confusion Matrix

Correct Classification	Clasified	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Berdasarkan informasi yang ditampilkan pada Tabel 2, evaluasi model dapat dilakukan menggunakan berbagai metrik seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Berikut penjelasan dari masing-masing metrik :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{5}$$

Persamaan (5) di mana *TP* (*True Positive*) adalah jumlah data positif yang terprediksi benar, *TN* (*True Negative*) adalah jumlah data negatif yang terprediksi benar, *FP* (*False Positive*) adalah jumlah data negatif yang salah diprediksi positif, dan *FN* (*False Negative*) adalah jumlah data positif yang salah diprediksi negatif.

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

Persamaan (6) di mana *precision* mengukur ketepatan model dalam memprediksi kelas positif dengan membandingkan jumlah prediksi positif yang benar terhadap seluruh prediksi positif.

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

Persamaan (7) di mana *recall* menunjukkan kemampuan model dalam mendeteksi seluruh data positif yang sebenarnya ada dalam dataset.

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall} \tag{8}$$

Persamaan (8) di mana *f1-score* memberikan keseimbangan antara ketepatan dan kelengkapan prediksi, terutama ketika data tidak seimbang.

3. HASIL DAN PEMBAHASAN

2.6 Pengumpulan Data

Data dalam penelitian ini diperoleh dari tweet masyarakat yang membahas topik “Tom Lembong” dalam bahasa Indonesia pada periode 1 Januari 2025 hingga 15 September 2025 di platform X. Pengambilan data dilakukan secara otomatis menggunakan teknik *crawling*, memanfaatkan kombinasi *tweet-harvest* berbasis Node.js dan pemrograman *Python*. Proses pengumpulan data difokuskan pada kata kunci tertentu, dengan target maksimum 2.000 entri. Namun, jumlah tweet yang berhasil dikumpulkan mencapai 1.802 entri karena keterbatasan ketersediaan data pada rentang waktu yang ditentukan. Seluruh data yang diperoleh kemudian diproses menggunakan *pandas DataFrame* untuk memudahkan pengelolaan, serta disimpan dalam format CSV dan Excel. Dataset ini terdiri dari 1.802 baris dan 15 kolom, siap untuk dianalisis lebih lanjut, termasuk identifikasi sentimen dan pola interaksi pengguna.

2.7 Pre-Processing

Sebelum dianalisis, data tweet diproses melalui *preprocessing* untuk membersihkan teks, menormalkan kata, melakukan tokenisasi, menghapus *stopwords*, dan *stemming*, seperti dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pre-Processing Data

Alur Proses	Hasil
Teks Asli	@barengwarga @celios_id Bagian menyalahkan diri sendiri itu yg berasa kesiksa banget ngalamin sendiri awal2 bisnis melambat 2-3 tahun lalu yg dipikirin cuma saya salah apa ya ini.... Yg bikin sadar ya pas podcast kasisolusi Tom Lembong saat beliau bilang jangan salahkan diri sendiri...
Cleaning Data	Bagian menyalahkan diri sendiri itu yg berasa kesiksa banget ngalamin sendiri awal bisnis melambat tahun lalu yg dipikirin cuma saya salah apa ya ini Yg bikin sadar ya pas podcast kasisolusi Tom Lembong saat beliau bilang jangan salahkan diri sendiri

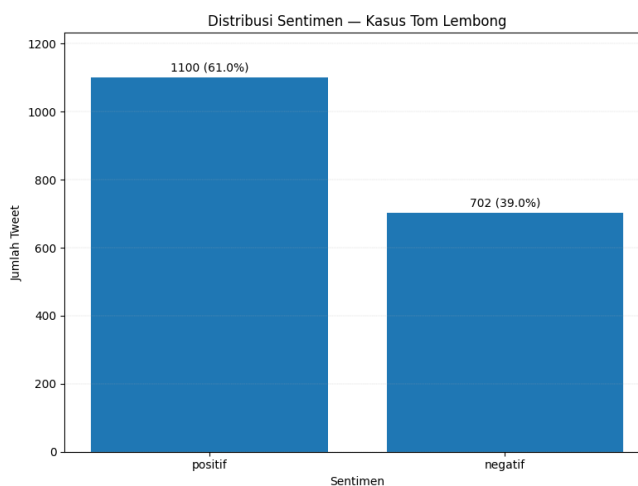
Alur Proses	Hasil
<i>Case Folding</i>	bagian menyalahkan diri sendiri itu yg berasa kesiksa banget ngalamin sendiri awal bisnis melambat tahun lalu yg dipikirin cuma saya salah apa ya ini yg bikin sadar ya pas podcast kasisolusi tom lembong saat beliau bilang jangan salahkan diri sendiri
Normalisasi Teks	bagian menyalahkan diri sendiri itu yang berasa kesiksa sangat ngalamin sendiri awal bisnis melambat tahun lalu yang dipikirkan hanya saya salah apa ya ini yang bikin sadar ya pas podcast kasisolusi tom lembong saat beliau bilang jangan salahkan diri sendiri
<i>Tokenizing</i>	bagian,menyalahkan,diri,sendiri,itu,yang,berasa,kesiksa,sangat,ngalamin,sendiri,awal,bisnis,mela mbat,tahun,lalu,yang,dipikirkan,hanya,saya,salah,apa,ya,ini,yang,bikin,sadar,ya,pas,podcast,kasi solusi,tom,lembong,saat,beliau,bilang,jangan,salahkan,diri,sendiri
<i>Stopword Removal</i>	menyalahkan,berasa,kesiksa,ngalamin,bisnis,melambat,dipikirkan,salah,ya,bikin,sadar,ya,pas,po dcast,kasisolusi,tom,lembong,beliau,bilang,salahkan
<i>Stemming</i>	salah,asa,siksa,ngalamin,bisnis,lambat,pikir,salah,ya,bikin,sadar,ya,pas,podcast,kasisolusi,tom,le mbong,beliau,bilang,salah
<i>Stemming String</i>	salah asa siksa ngalamin bisnis lambat pikir salah ya bikin sadar ya pas podcast kasisolusi tom lembong beliau bilang salah

2.8 Pengolahan Data

Pengolahan Data meliputi Pelabelan Data menggunakan *Lexicon-Based*, eksplorasi data, *Feature Extraction* menggunakan TF-IDF dan *Split* Dataset.

3.3.1 Pelabelan Data

Setelah data tweet melalui tahap *preprocessing*, setiap tweet diberi skor sentimen menggunakan metode *lexicon* dengan menjumlahkan bobot kata yang terdapat pada kamus sentimen. Berdasarkan skor tersebut, tweet diberi label positif jika skor ≥ 0 dan negatif jika skor < 0 . Hasil Pelabelan ditampilkan pada Gambar 2.



Gambar 2. Hasil Pelabelan Data

Gambar 2 menunjukkan bahwa sentimen publik terhadap kasus Tom Lembong didominasi oleh opini positif, dengan 61,0% (1100 tweet) bersentimen positif dan 39,0% (702 tweet) bersentimen negatif. Hal ini mengindikasikan bahwa mayoritas warganet memberikan tanggapan yang cenderung mendukung atau berpandangan baik terhadap Tom Lembong dalam kasus tersebut.

Tabel 4. Hasil Pelabelan Data

Teks	Skor Sentimen	Label Sentimen
salah asa siksa ngalamin bisnis lambat pikir salah ya bikin sadar ya pas podcast kasisolusi tom lembong beliau bilang salah	4	Positif
dasar dakwa jagung sumber hukumonline kompas impor gula tom lembong rugi negara rp miliar kaya usaha swasta instansi pemerintah untung jaksa tom klaim pt ppi bumh	-2	Negatif
dasar dakwa jagung usaha kaya impor gula tom lembong total rp miliar tuju impor koordinasi menteri libat kerjasama pt ppi inkopkar dll pt angels products rp miliar pt makassar tene	13	Positif
case tom lembong lhi dll ngeri terap hukum mati negara mental tegak hukum korporasi kaya tom lembong tahap lidi	6	Positif
	2	Positif

Tabel 4 tersebut menampilkan contoh hasil pelabelan sentimen dari beberapa teks tweet terkait kasus Tom Lembong. Setiap teks diklasifikasikan ke dalam kategori Positif atau Negatif berdasarkan skor sentimen yang dihasilkan.

3.3.2 Eksplorasi Data

Untuk mengidentifikasi kata-kata dominan dalam tweet terkait "Tom Lembong", dilakukan visualisasi menggunakan WordCloud berdasarkan seluruh data maupun label sentimen. Visualisasi wordcloud sentimen positif dan sentimen negatif dapat dilihat pada Gambar 3 dan Gambar 4.



Gambar 3. Wordcloud Sentimen Positif



Gambar 4. Wordcloud Sentimen Negatif

Gambar 3 merupakan visualisasi Wordcloud Sentimen Positif yang menampilkan kata-kata yang paling sering muncul pada tweet dengan sentimen positif terkait kasus Tom Lembong. Kata “Tom Lembong” mendominasi bersama istilah lain seperti *referral*, *pasuk*, dan *adil*, yang menunjukkan adanya konteks percakapan positif seputar aktivitas profesional maupun keadilan. Sedangkan Gambar 4 merupakan *Wordcloud* Sentimen Negatif, yang memperlihatkan kata-kata dominan pada tweet bernada negatif seperti *korupsi*, *rugi negara*, *bukti*, dan *penjara*. Kata-kata tersebut menunjukkan adanya opini publik yang mengaitkan Tom Lembong dengan isu-isu hukum dan kerugian negara, mencerminkan persepsi negatif terhadap kasus yang sedang dibahas.

3.3.3 Feature Extraction

Untuk menyiapkan data teks sebagai fitur, digunakan TF-IDF (*Term Frequency–Inverse Document Frequency*). Metode ini menilai kepentingan kata dalam korpus tweet, sehingga kata dengan bobot TF-IDF tinggi menjadi fitur utama untuk analisis sentimen. Ringkasan TF-IDF dapat dilihat pada Tabel 5.

Tabel 5. Hasil TF-IDF

No	Term	TF-IDF	TF
0	tom	122.43580	1505
1	lembong	120.90015	1468
2	nadiem	45.62798	301
3	ya	41.67108	242
4	kayak	33.71697	132
...
3740	menguru	0.14706	1
3741	soemitro	0.14706	1

No	Term	TF-IDF	TF
3742	kroso	0.14706	1
3743	percoyo	0.14706	1

3.3.4 Split Dataset

Untuk mengetahui proporsi data yang digunakan pada tahap pelatihan dan pengujian model, dilakukan pembagian dataset berdasarkan label sentimen. Dataset dibagi menjadi dua subset utama, yaitu data latih (*training data*) dan data uji (*testing data*), yang masing-masing mencakup dua kategori sentimen, yaitu positif dan negatif. Rincian jumlah data pada setiap kategori dapat dilihat pada Tabel 6.

Tabel 6. Hasil Split Data

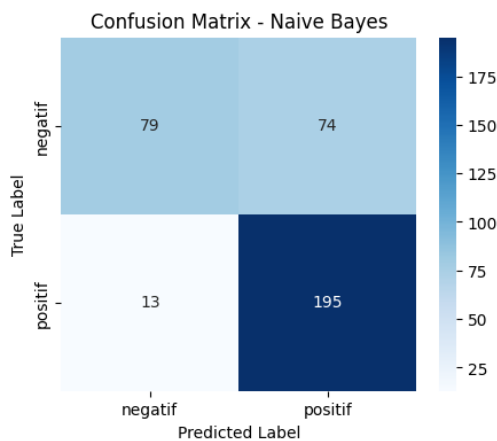
Dataset	Label	Jumlah
Latih	Positif	892
Latih	Negatif	549
Uji	Positif	208
Uji	Negatif	153

3.4 Hasil dan Analisa

Hasil dan Analisa pada dataset “Tom Lembong” hasil crawl dari X menggunakan *Confusion Matrix* (CM) yang terdiri dari *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN) untuk menilai ketepatan model dalam klasifikasi sentimen positif dan negatif.

3.4.1 Naïve Bayes

Evaluasi performa model dilakukan menggunakan *Confusion Matrix* untuk melihat tingkat ketepatan prediksi pada tiap kelas sentimen, *Naïve Bayes* biasanya menggunakan *Multinomial Naïve Bayes* (*MultinomialNB*). Hasil prediksi model *Naïve Bayes* terhadap data uji dapat dilihat pada Gambar 5.



Gambar 5. Confusion Matrix - Naïve Bayes

Gambar 5 menampilkan *Confusion Matrix* model *Naïve Bayes* dengan hasil TP = 195, TN = 79, FP = 74, dan FN = 13, yang menunjukkan model lebih akurat dalam mengenali sentimen positif dibandingkan negatif.

```

=== Naive Bayes ===
Akurasi: 0.7590027700831025
precision    recall  f1-score   support

negatif      0.86    0.52    0.64     153
positif      0.72    0.94    0.82     208

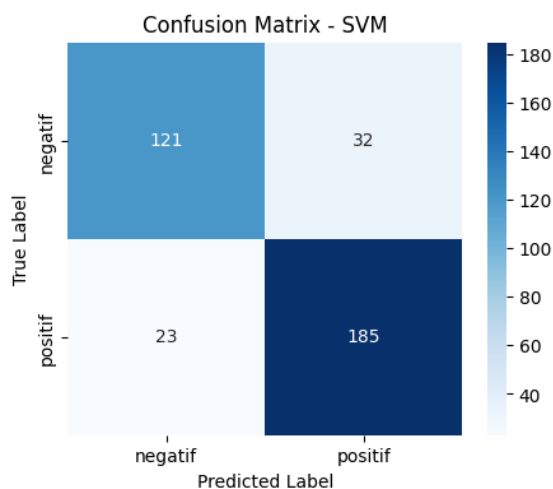
accuracy          0.76     361
macro avg         0.79    0.73    0.73     361
weighted avg     0.78    0.76    0.74     361
    
```

Gambar 6. Evaluasi Model Naïve Bayes

Berdasarkan Gambar 6, model *Naïve Bayes* memperoleh akurasi sebesar 0.76 dengan nilai *precision* 0.86 untuk kelas negatif dan 0.72 untuk kelas positif. Nilai *recall* menunjukkan model lebih baik dalam mengenali sentimen positif (0.94) dibanding negatif (0.52). Secara keseluruhan, performa model cukup baik, meskipun masih terdapat ketidakseimbangan dalam mendeteksi kelas negatif.

3.4.2 Support Vector Machine

Model SVM menggunakan *LinearSVC* dan dievaluasi dengan *Confusion Matrix* untuk menilai akurasi tiap kelas sentimen. Hasil prediksi ditunjukkan pada Gambar 7.



Gambar 7. Confusion Matrix - Support Vector Machine

Berdasarkan Gambar 7, model SVM berhasil memprediksi dengan benar 121 data negatif (TN) dan 185 data positif (TP), sedangkan kesalahan prediksi terjadi pada 32 data negatif (FP) dan 23 data positif (FN).

```

=== SVM ===
Akurasi: 0.8476454293628809
precision  recall  f1-score  support
negatif    0.84    0.79    0.81     153
positif    0.85    0.89    0.87     208

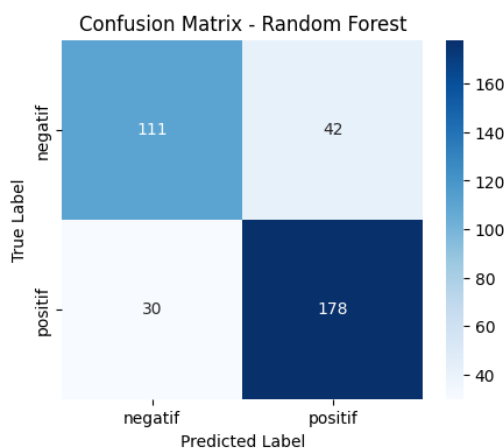
accuracy              0.85     361
macro avg             0.85    0.84    0.84     361
weighted avg          0.85    0.85    0.85     361
    
```

Gambar 8. Evaluasi Model Support Vector Machine

Berdasarkan hasil evaluasi pada Gambar 8, model SVM mencapai akurasi sebesar 0.85, dengan *precision* dan *recall* yang seimbang pada kedua kelas sentimen. Kinerja terbaik ditunjukkan pada kelas positif dengan *precision* 0.85, *recall* 0.89, dan *F1-score* 0.87, menunjukkan bahwa model mampu mengenali sentimen positif dengan sangat baik serta menjaga keseimbangan performa secara keseluruhan.

3.4.3 Random Forest

Model Random Forest dievaluasi menggunakan *Confusion Matrix* untuk menilai akurasi pada setiap kelas sentimen. Hasil prediksi model Random Forest ditunjukkan pada Gambar 9.



Gambar 9. Confusion Matrix - Random Forest

Berdasarkan Gambar 9, model Random Forest memiliki hasil: TP = 178, TN = 111, FP = 42, dan FN = 30, menunjukkan kinerja yang cukup baik meskipun masih ada beberapa kesalahan prediksi pada kelas negatif.

```

=== Random Forest ===
Akurasi: 0.8005540166204986
precision  recall  f1-score  support
negatif    0.79    0.73    0.76    153
positif    0.81    0.86    0.83    208

accuracy
macro avg  0.80    0.79    0.79    361
weighted avg 0.80    0.80    0.80    361
    
```

Gambar 10. Evaluasi Model Random Forest

Berdasarkan hasil evaluasi pada Gambar 10, model *Random Forest* memperoleh akurasi sebesar 0.80, menunjukkan performa yang cukup baik dalam klasifikasi sentimen. Model ini memiliki *precision* 0.81 dan *recall* 0.86 pada kelas positif, serta *precision* 0.79 dan *recall* 0.73 pada kelas negatif. Hal ini menunjukkan bahwa model lebih optimal dalam mengenali sentimen positif dibandingkan negatif.

3.5 Evaluasi Perbandingan Model

Berikut merupakan hasil rincian perbandingan mengenai pengujian algoritma tersebut pada Tabel 7.

Tabel 7. Hasil Perbandingan Confusion Matrix Model

Model	True Positive	True Negative	False Positive	False Negative
Naïve Bayes	195	79	74	13
Support Vector Machine	185	121	32	23
Random Forest	178	111	42	30

Berdasarkan hasil pada Tabel 7, model *Support Vector Machine* (SVM) menunjukkan performa terbaik dengan *True Positive* (185) dan *True Negative* (121) tertinggi, serta *False Positive* (32) dan *False Negative* (23) paling rendah dibandingkan model lain. Model *Random Forest* menempati posisi kedua dengan hasil yang cukup seimbang (TP 178, TN 111), sedangkan *Naïve Bayes* memiliki akurasi terendah karena menghasilkan *False Positive* (74) paling tinggi. Secara keseluruhan, SVM adalah model paling akurat dan andal dalam klasifikasi sentimen pada dataset ini.

Tabel 8. Hasil Perbandingan Model

Model	Class	Accuracy	Precision	Recall	F1-score
Naïve Bayes	Negative	0.76	0.86	0.52	0.64
	Positive		0.72	0.94	0.82
Support Vector Machine	Negative	0.84	0.84	0.79	0.81
	Positive		0.85	0.89	0.87
Random Forest	Negative	0.80	0.79	0.73	0.76
	Positive		0.81	0.86	0.83

Berdasarkan hasil perbandingan model pada Tabel 8, algoritma *Support Vector Machine* (SVM) menunjukkan kinerja terbaik dengan akurasi tertinggi sebesar 0.84 serta *F1-score* 0.87 pada kelas positif. Model ini mampu menjaga keseimbangan antara *precision* dan *recall* pada kedua kelas sentimen. Sementara itu, *Random Forest* menempati posisi kedua dengan akurasi 0.80, dan *Naïve Bayes* memiliki akurasi terendah (0.76) serta *recall* rendah pada kelas negatif.

Hasil ini menunjukkan bahwa model SVM memiliki kemampuan generalisasi yang baik dalam merepresentasikan pola sentimen pada dataset penelitian ini, bahwa proses evaluasi dilakukan berdasarkan label yang dihasilkan oleh metode *lexicon-based*, sehingga hasil akurasi merefleksikan sejauh mana model *Machine Learning* mampu meniru pola klasifikasi dari pendekatan tersebut. Dengan demikian, akurasi tinggi yang dicapai oleh SVM menunjukkan konsistensi dan kesesuaian model terhadap pola penilaian yang telah ditetapkan oleh sistem *lexicon-based*. Meskipun demikian, interpretasi hasil tetap mempertimbangkan bahwa proporsi sentimen positif yang mendominasi (61%) bergantung pada karakteristik dan cakupan kamus sentimen yang digunakan. Oleh karena itu, temuan ini diinterpretasikan bahwa SVM merupakan model paling andal dalam mengklasifikasikan sentimen sesuai standar *lexicon-based*, sekaligus memiliki potensi kuat untuk digunakan pada sistem analisis sentimen dengan dataset yang lebih luas dan beranotasi manual di penelitian selanjutnya.

4. KESIMPULAN

Berdasarkan hasil analisis terhadap data tweet pengguna media sosial X mengenai kasus dugaan korupsi impor gula yang melibatkan Tom Lembong, diperoleh bahwa opini publik cenderung bersifat positif. Dari total 1.802 tweet yang

dianalisis, sebanyak 1.100 tweet (61%) dikategorikan sebagai sentimen positif dan 702 tweet (39%) sebagai sentimen negatif. Temuan ini menunjukkan dominasi opini positif dalam percakapan publik mengenai kasus tersebut. Namun, dalam isu hukum, hasil ini perlu diinterpretasikan secara kritis dan *non-literal*, karena sentimen positif tidak selalu merefleksikan dukungan murni terhadap terlapor. Opini positif yang dominan dapat mengindikasikan bahwa persepsi publik lebih dipengaruhi oleh reputasi pribadi Tom Lembong, posisi politiknya, atau bentuk kritik terhadap lembaga penegak hukum dan perhitungan kerugian negara, bukan semata-mata pembelaan terhadap dugaan tindak pidana korupsi. Proses pelabelan data dilakukan menggunakan pendekatan *lexicon-based*, yang berfungsi menentukan polaritas sentimen sebagai acuan dalam pelatihan model *machine learning*. Hasil pengujian terhadap tiga algoritma—*Naïve Bayes*, *Support Vector Machine (SVM)*, dan *Random Forest*—menunjukkan bahwa SVM memiliki performa terbaik dengan akurasi 84%. Nilai tersebut menegaskan kemampuan SVM dalam mengenali pola klasifikasi sentimen dengan baik serta menjaga keseimbangan antara ketepatan dan kelengkapan prediksi. Dengan demikian, penelitian ini menunjukkan bahwa SVM efektif digunakan untuk analisis sentimen publik terhadap isu-isu hukum di Indonesia. Meskipun demikian, penelitian ini memiliki keterbatasan karena proses pelabelan masih mengandalkan metode *lexicon-based*, yang sensitif terhadap konteks dan makna implisit seperti *sarkasme* atau ironi. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan anotasi manual oleh manusia (*human annotation*) guna meningkatkan validitas data dan menangkap nuansa linguistik yang lebih kompleks. Selain itu, perluasan sumber data, rentang waktu pengumpulan, penerapan metode penyeimbangan data seperti SMOTE, serta penerapan pendekatan *deep learning* seperti LSTM atau BERT diharapkan dapat menghasilkan analisis sentimen yang lebih akurat dan kontekstual terhadap dinamika opini publik.

REFERENCES

- [1] M. Rizki, “Tom Lembong Didakwa Rugikan Negara Rp578 Miliar di Kasus Dugaan Korupsi Importasi Gula,” hukumonline.com. [Online]. Available: <https://www.hukumonline.com/berita/a/tom-lembong-didakwa-rugikan-negara-rp578-miliar-di-kasus-dugaan-korupsi-importasi-gula-lt67c96d4063fde/>
- [2] Rizky, “Jaksa Sebut Rp578 Miliar, Hakim Putuskan Hanya Rp194 Miliar, Vonis Tom Lembong Bikin Geger,” tempo.co. [Online]. Available: <https://www.tempo.co/hukum/vonis-tom-lembong-4-5-tahun-penjara-angka-kerugian-negara-versi-hakim-dan-jaksa-berbeda-2048843>
- [3] A. Sadat, “Akademisi Kritik Hitungan Kerugian Negara dalam Kasus Impor Gula Tom Lembong,” viva.co.id. [Online]. Available: <https://www.viva.co.id/berita/nasional/1834004-akademisi-kritik-hitungan-kerugian-negara-dalam-kasus-import-gula-tom-lembong>
- [4] S. Ni’am and N. Akbar, “Divonis 4,5 Tahun Penjara, Hakim: Tom Lembong Tak Nikmati Hasil Divonis 4,5 Tahun Penjara, Hakim: Tom Lembong Tak Nikmati Hasil Korupsi,” kompas.com. [Online]. Available: <https://nasional.kompas.com/read/2025/07/18/18462411/divonis-45-tahun-penjara-hakim-tom-lembong-tak-nikmati-hasil-korupsi>
- [5] K. Pramayasa, I. Md, D. Maysanjaya, G. Ayu, and A. Diatri Indradewi, “Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE,” *Science and Information Technology (SINTECH)*, vol. 6, Aug. 2023, doi: <https://doi.org/10.31598>.
- [6] D. Prasetyoningrum and P. N. Andono, “Perbandingan Model Machine Learning dalam Analisis Sentimen Pada Kasus Monkeypox di Media Sosial X,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, pp. 2005–2014, Dec. 2024, doi: [10.47065/bits.v6i3.6447](https://doi.org/10.47065/bits.v6i3.6447).
- [7] A. H. A. Zharifa and E. I. H. Ujianto, “Analisis Sentimen Publik di Twitter Pasca Debat Kelima Pilpres 2024 dengan Naive Bayes,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 754–763, Dec. 2024, doi: [10.29408/edumatic.v8i2.28048](https://doi.org/10.29408/edumatic.v8i2.28048).
- [8] D. A. Fitri and D. Damayanti, “Komparasi Algoritma Random Forest Classifier Dan Support Vector Machine Untuk Sentimen Masyarakat Terhadap Pinjaman Online Di Media Sosial,” *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 4, pp. 2018–2029, Nov. 2024, doi: [10.29100/jipi.v9i4.5608](https://doi.org/10.29100/jipi.v9i4.5608).
- [9] E. Triningsih, M. Afdal, I. Permana, and N. Evrilyan Rozanda, “Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 4, pp. 2240–2250, Mar. 2025, doi: [10.47065/bits.v6i4.6534](https://doi.org/10.47065/bits.v6i4.6534).
- [10] E. Danuarta and D. Alita, “Perbandingan Algoritma Naïve Bayes dan Random Forest untuk Melakukan Analisis Sentimen Cyberbullying Generasi Z Pada Twitter,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 4, Jul. 2025, doi: [10.47065/bits.v6i4.6909](https://doi.org/10.47065/bits.v6i4.6909).
- [11] N. Shabrina Nasution, I. Permana, F. Nur Salisah, M. Afdal, and M. Megawati, “Analisis Sentimen Masyarakat Terhadap Kebijakan IKN Pada Periode Jokowi dan Prabowo Menggunakan Algoritma NBC, SVM, dan K-NN,” *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 1, 2025, doi: [10.47065/bits.v7i1.7276](https://doi.org/10.47065/bits.v7i1.7276).
- [12] M. Febrian As Shidiq, D. Alita, and L. Ratu Kec Kedaton Kota Bandar Lampung, “Analisis Sentimen Masyarakat Terhadap Kasus Judi Online Menggunakan Data dari Media Sosial X Pendekatan Naive Bayes Dan SVM,” *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, 2025, doi: [10.47080/simika.v8i1.3624](https://doi.org/10.47080/simika.v8i1.3624).
- [13] A. Setiawan and R. R. Suryono, “Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 183–192, Jun. 2024, doi: [10.29408/edumatic.v8i1.25667](https://doi.org/10.29408/edumatic.v8i1.25667).
- [14] F. Arrazak, M. Afdal, R. Novita, and M. Megawati, “Analisis Sentimen Terhadap Pemain Naturalisasi dan Lokal Tim Nasional Sepakbola Indonesia Menggunakan Support Vector Machine,” *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 2, Sep. 2025, doi: [10.47065/bits.v7i2.7471](https://doi.org/10.47065/bits.v7i2.7471).



- [15] M. Daffa, A. Fahreza, A. Luthfiarta, M. Rafid, M. Indrawan, and A. Nugraha, “Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z,” *Journal Of Applied Computer Science And Technology (JACOST)*, vol. 5, no. 1, pp. 2723–1453, 2024, doi: 10.52158/jacost.715.
- [16] E. Arya Pranata, F. Budiman, and D. Kurniawan, “Analisis Sentimen Ulasan Mobile JKN pada Playstore dengan Perbandingan Akurasi Algoritma Naïve Bayes dan SVM,” *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 1, Jun. 2025, doi: 10.47065/bits.v7i1.7334.
- [17] N. Van Robert Jhosefhin *et al.*, “Analisis Sentimen Crawling Data dari Sosial Media X tentang Gaza Menggunakan Metode SVM dan Decision Tree,” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi (JIMIK)*, vol. 6, no. 1, 2025, doi: 10.3587.
- [18] M. Irwanda, M. Afdal, R. Novita, and Z. Zarnelly, “Analisis Sentimen Masyarakat Terhadap Liga Indonesia Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector Machine Pada Platform X dan YouTube,” *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 1, 2025, doi: 10.47065/bits.v7i1.7294.
- [19] E. Yuniar and N. Hendrastuty, “Perbandingan Metode Naïve Bayes, Random Forest dan SVM Untuk Analisis Sentimen Pada Twitter Tentang Kenaikan Gaji Guru,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 4, Mar. 2025, doi: 10.47065/bits.v6i4.6970.
- [20] Z. Alfian, M. Afdal, R. Novita, and Z. Zarnelly, “Penerapan Support Vector Machine untuk Analisis Sentimen Pengguna X terhadap IndiHome, Biznet, dan Starlink,” *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 2, Sep. 2025, doi: 10.47065/bits.v7i2.7429.
- [21] M. G. Andriawan and T. Ernawati, “Penggunaan Algoritma Naïve Bayes dan Support Vector Machine untuk Analisis Sentimen Konflik Palestina Dan Israel Pada Platform X,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4943.
- [22] A. Safira and F. N. Hasan, “Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier,” *Jurnal Sistem Informasi*, vol. 5, no. 1, Jan. 2023, doi: 10.31849/zn.v5i1.12856.
- [23] J. Khatib Sulaiman, H. Margono, F. Nur Aprillia, and U. Airlangga, “Analisis Sentimen Komentar Childfree di Aplikasi X Menggunakan Naïve Bayes,” *Indonesian Journal of Computer Science*, vol. 13, no. 3, Apr. 2024, doi: 10.33022/ijcs.v13i3.3883.
- [24] N. Hadi and D. Sugiarto, “Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 1, pp. 37–49, Jan. 2025, doi: 10.30591/jpit.v10i1.7106.
- [25] R. Raihandika and R. Suryono, “Perbandingan Algoritma Random Forest, KNN, SVM Untuk Analisis Sentimen Pengalaman Belanja ThriftDi X,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, No. 4, 2025, doi: 10.47065/bits.v6i4.6797.
- [26] G. P. Insany, I. L. Kharisma, and R. Rosmawati, “Penerapan Algoritma Random Forest untuk Menganalisis Ulasan Aplikasi Spotify pada Google Play,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 369–378, Dec. 2024, doi: 10.29408/edumatic.v8i2.26394.
- [27] E. Darwisah Harahap and R. Kurniawan, “Analisis Sentimen Komentar Terhadap Kebijakan Pemerintah Mengenai Tabungan Perumahan Rakyat (TAPERA) Pada Aplikasi X Menggunakan Metode Naïve Bayes,” *Jurnal Teknik Informatika Unika ST. Thomas (JTIUST)*, vol. 9, no. 1, pp. 2657–1501, Jun. 2024, doi: 10.54367/jtiust.v9i1.3911.