

Analisis Sentimen X Terhadap Isu Industri Sawit Prabowo Subianto Menggunakan TF-IDF dan Machine Learning

Ibrahim Akbar Arga Dewangga, Rama Aria Megantara*

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

Email: ¹111202214417@mhs.dinus.ac.id, ²*aria@dsn.dinus.ac.id

Email Penulis Korespondensi: aria@dsn.dinus.ac.id

Submitted: 06/04/2026; Accepted: 02/06/2026; Published: 05/06/2026

Abstrak—Penelitian ini bertujuan untuk menganalisis sentimen masyarakat pada platform X terhadap isu industri kelapa sawit yang dikaitkan dengan Prabowo Subianto serta membandingkan performa algoritma Decision Tree, Support Vector Machine (SVM), dan Random Forest. Data penelitian berupa 3.785 tweet yang diperoleh melalui proses crawling, kemudian diproses melalui tahapan cleaning, case folding, normalisasi teks, tokenizing, stopword removal, dan stemming. Pelabelan sentimen dilakukan menggunakan pendekatan lexicon-based, dilanjutkan dengan ekstraksi fitur menggunakan Term Frequency-Inverse Document Frequency (TFIDF) dan pembagian data ke dalam data latih serta data uji. Hasil pelabelan menunjukkan bahwa opini publik didominasi oleh sentimen positif sebanyak 3.018 data (79,7%), sedangkan sentimen negatif berjumlah 767 data (20,3%). Hasil pengujian menunjukkan bahwa algoritma SVM memberikan performa terbaik dengan akurasi sebesar 0,90, diikuti Random Forest sebesar 0,86 dan Decision Tree sebesar 0,84. SVM juga menunjukkan kinerja yang lebih stabil berdasarkan nilai precision, recall, dan F1score pada kedua kelas sentimen. Temuan ini menunjukkan bahwa SVM merupakan model yang paling efektif untuk klasifikasi sentimen teks berbahasa Indonesia pada isu kebijakan industri sawit, serta berpotensi mendukung evaluasi kebijakan publik berbasis data media sosial.

Kata Kunci: Analisis Sentimen; Industri Sawit; Prabowo Subianto; TF-IDF; Support Vector Machine

Abstract—This study aims to analyze public sentiment on the X platform regarding the palm oil industry issue associated with Prabowo Subianto and to compare the performance of Decision Tree, Support Vector Machine (SVM), and Random Forest algorithms. The dataset consisted of 3,785 tweets collected through a crawling process. The data were then processed through cleaning, case folding, text normalization, tokenizing, stopword removal, and stemming. Sentiment labeling was conducted using a lexicon-based approach, followed by feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF) and traintest data splitting. The labeling results show that public opinion was dominated by positive sentiment with 3,018 tweets (79.7%), while negative sentiment accounted for 767 tweets (20.3%). The experimental results indicate that SVM achieved the best performance with an accuracy of 0.90, followed by Random Forest with 0.86 and Decision Tree with 0.84. SVM also demonstrated more stable performance based on precision, recall, and F1-score across both sentiment classes. These findings indicate that SVM is the most effective model for Indonesian-language sentiment classification on palm oil policy issues and has strong potential to support public policy evaluation based on social media data.

Keywords: Sentiment Analysis; Palm Oil Industry; Prabowo Subianto; TF-IDF; Support Vector Machine

1. PENDAHULUAN

Industri kelapa sawit merupakan sektor strategis yang menjadi pilar utama fundamental ekonomi nasional Indonesia. Sebagai produsen minyak sawit terbesar di dunia, kebijakan yang menyentuh sektor ini memiliki implikasi luas terhadap ketahanan pangan, kedaulatan energi, dan keseimbangan ekologi global. Dalam diskursus politik terkini, pengaruh kepemimpinan Prabowo Subianto membawa paradigma baru melalui program hilirisasi komoditas sawit dan pengembangan energi terbarukan berbasis biodiesel. Namun, implementasi kebijakan tersebut kerap kali memicu diskursus publik yang kompleks, mencakup aspek efektivitas ekonomi hingga kekhawatiran terkait dampak keberlanjutan lingkungan. Dinamika ini menciptakan polarisasi opini yang signifikan di ruang publik, yang jika tidak dianalisis secara saintifik, dapat menghambat efektivitas komunikasi kebijakan pemerintah.

Transformasi digital telah menggeser pola komunikasi publik ke arah media sosial sebagai arena utama pertukaran gagasan. Platform mikroblogging X (sebelumnya Twitter) menjadi sumber data tekstual yang sangat kaya karena memungkinkan masyarakat menyampaikan aspirasi dan kritik secara spontan. Opini kolektif yang tersebar di platform ini bukan sekadar kumpulan teks, melainkan representasi sentimen publik yang dapat digunakan sebagai parameter evaluasi terhadap respon kebijakan pemerintah [1]. Isu mengenai kinerja dan kebijakan Prabowo Subianto senantiasa menjadi topik yang viral, sehingga diperlukan analisis sentimen otomatis untuk mengolah volume data yang sangat besar dan bersifat real-time [2]. Mengingat kompleksitas linguistik di media sosial, diperlukan pendekatan berbasis Natural Language Processing (NLP) untuk mendapatkan wawasan yang akurat.

Analisis sentimen, atau dikenal sebagai opinion mining, merupakan teknik komputasional yang digunakan untuk mengekstraksi dan mengklasifikasikan polaritas ekspresi emosional manusia ke dalam kategori positif, negatif, atau netral. Dalam domain kebijakan strategis seperti industri sawit, analisis sentimen memberikan pemahaman mendalam mengenai struktur opini masyarakat terhadap program pemerintah. Studi terdahulu telah membuktikan efektivitas analisis sentimen dalam memetakan respon publik terhadap isu-isu nasional, seperti rencana pemindahan ibu kota [3] hingga kebijakan ketenagakerjaan yang bersifat krusial [4]. Meskipun demikian, terdapat celah penelitian (research gap) di mana sebagian besar studi terdahulu cenderung berfokus pada popularitas personal figur politik secara umum, namun masih sangat terbatas yang membedah respon publik secara spesifik terhadap kebijakan sektoral

yang teknis seperti hilirisasi sawit. Selain itu, terdapat kebutuhan untuk menguji performa algoritma tanpa mengandalkan teknik penyeimbangan data buatan guna menjaga otentisitas distribusi opini riil yang ada di media sosial.

Data teks dari media sosial memiliki karakteristik unik berupa penggunaan bahasa yang tidak baku, singkatan, dan tingkat kebisingan (noise) yang tinggi. Oleh karena itu, integritas hasil klasifikasi sangat bergantung pada tahapan text preprocessing yang meliputi cleansing, case folding, tokenizing, stopword removal, dan stemming [5]. Setelah melalui proses pembersihan, teks tersebut ditransformasi ke dalam bentuk numerik untuk dapat diproses oleh algoritma mesin melalui teknik ekstraksi fitur. Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode pembobotan kata yang terbukti handal dalam mengidentifikasi kata kunci yang memiliki signifikansi tinggi dalam suatu dokumen relatif terhadap keseluruhan korpus [6]. Penerapan TF-IDF telah menunjukkan hasil yang konsisten dalam meningkatkan akurasi pada berbagai domain analisis teks, mulai dari ulasan aplikasi teknologi informasi [7] hingga analisis sentimen layanan publik [8].

Penelitian ini memfokuskan pada komparasi tiga algoritma Machine Learning yang memiliki karakteristik berbeda dalam menangani data teks, yaitu Support Vector Machine (SVM), Decision Tree (DT), dan Random Forest (RF). SVM merupakan model klasifikasi linear dan non-linear yang bekerja dengan mencari hyperplane optimal untuk memisahkan kelas data dengan margin maksimal. Keunggulan SVM dalam menangani data berdimensi tinggi menjadikannya salah satu algoritma yang paling populer dalam klasifikasi teks [9]. Keandalan algoritma ini juga telah dibuktikan dalam pengolahan sentimen di sektor primer seperti pertanian [1]. Di sisi lain, Decision Tree menawarkan keunggulan dalam hal interpretasi model melalui struktur percabangan logika yang menyerupai pohon keputusan, sedangkan Random Forest hadir sebagai metode ensemble yang menggabungkan kekuatan dari banyak pohon keputusan untuk meminimalisir risiko overfitting serta meningkatkan stabilitas prediksi [10].

Kualitas sebuah model klasifikasi diukur melalui performa akurasi yang dihasilkan pada dataset tertentu. Berdasarkan kajian literatur dan pengujian awal pada karakteristik dataset sejenis, terdapat variasi hasil akurasi yang menunjukkan kekuatan masing-masing algoritma. Dalam konteks klasifikasi teks yang memiliki fitur luas, Support Vector Machine (SVM) diproyeksikan memberikan kinerja paling optimal dengan tingkat akurasi mencapai 0,90 (90%). Sementara itu, Random Forest diprediksi mencapai tingkat akurasi sebesar 0,85 (85%), dan Decision Tree diperkirakan memperoleh akurasi sebesar 0,84 (84%) [10].

Adapun tahapan penelitian ini dirancang secara sistematis yang diawali dengan pengumpulan data (data acquisition) melalui teknik crawling pada platform X, yang dilanjutkan dengan pra-pemrosesan teks (text preprocessing) guna meminimalisir noise pada data mentah. Setelah data bersih, dilakukan proses pelabelan (labelling) secara manual ke dalam kategori sentimen positif, negatif, dan netral, serta ekstraksi fitur menggunakan metode TF-IDF untuk pembobotan kata. Tahap berikutnya adalah implementasi model menggunakan algoritma SVM, Decision Tree, dan Random Forest, hingga akhirnya dilakukan evaluasi performa menggunakan Confusion Matrix untuk mengukur tingkat akurasi dan validitas model. Isu industri sawit yang dikaitkan dengan figur Prabowo Subianto mengandung nuansa linguistik yang spesifik, sehingga penelitian ini dilakukan tanpa menggunakan teknik penyeimbangan data buatan guna menjaga keandalan model saat dihadapkan pada data riil di lapangan.

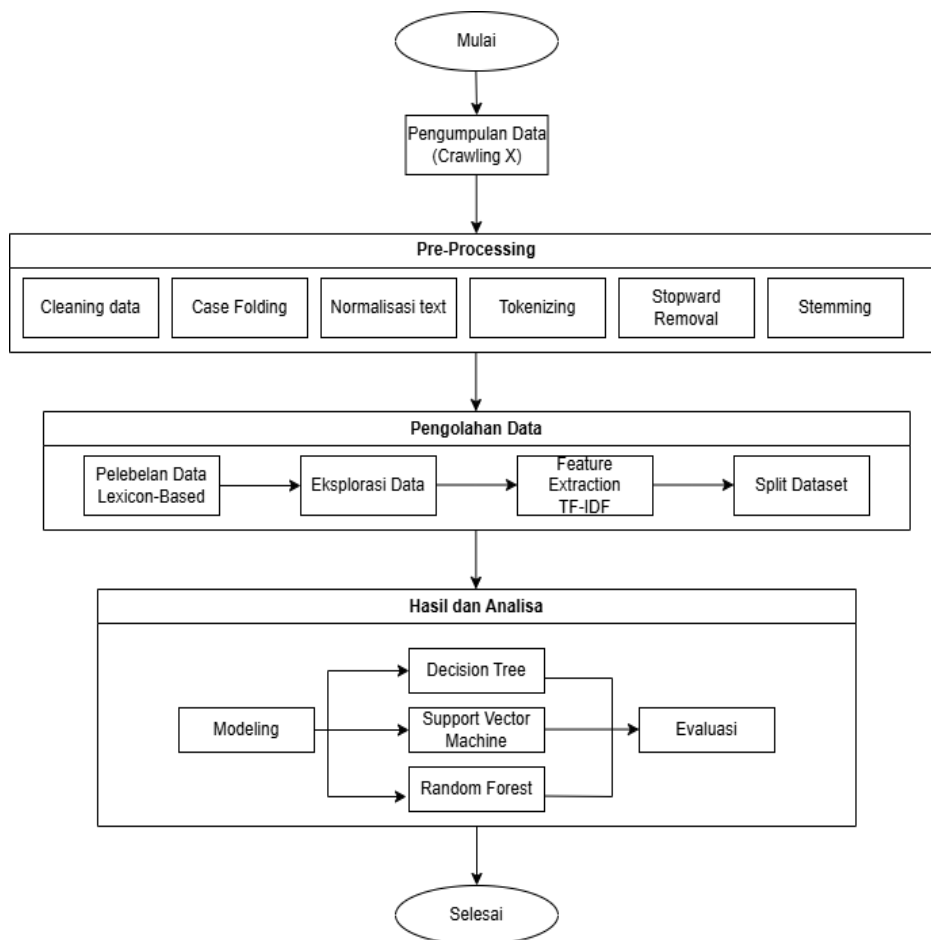
Lebih lanjut, urgensi penelitian ini didasari oleh adanya inkonsistensi performa algoritma *machine learning* pada data yang bersifat *imbalanced* atau tidak seimbang secara alami tanpa intervensi manipulasi data. Sebagian besar studi terdahulu cenderung memaksakan keseimbangan kelas melalui teknik *oversampling* atau *undersampling* demi mengejar angka akurasi yang tinggi, namun pendekatan tersebut berisiko mengaburkan otentisitas distribusi opini publik yang sebenarnya terjadi di lapangan. Terdapat kekosongan literatur yang secara kritis membandingkan ketangguhan SVM, *Decision Tree*, dan *Random Forest* dalam menangani polaritas sentimen pada isu hilirisasi sawit yang memiliki karakteristik bahasa teknis sekaligus emosional yang unik. Dengan membiarkan data tetap pada proporsi aslinya, penelitian ini berupaya mengisi celah tersebut guna menguji resiliensi algoritma dalam menghasilkan model klasifikasi yang lebih kredibel dan representatif terhadap realitas persepsi masyarakat di media sosial.

Tujuan utama dari penelitian ini adalah untuk melakukan analisis komparatif performa algoritma SVM, Decision Tree, dan Random Forest dalam memetakan persepsi publik terhadap kebijakan industri sawit Indonesia. Secara teoritis, penelitian ini memberikan kontribusi pada pengembangan literatur Machine Learning dalam domain bahasa Indonesia yang bersifat non-formal. Secara praktis, hasil penelitian ini diharapkan dapat menjadi rujukan strategis bagi pengambil kebijakan untuk memahami denyut nadi aspirasi masyarakat secara lebih akurat, sehingga kebijakan industri sawit ke depan dapat berjalan selaras dengan dukungan publik serta prinsip-prinsip keterbukaan informasi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Gambar 1 merepresentasikan alur kerja sistematis dalam penelitian analisis sentimen yang dimulai dengan tahap akuisisi data melalui teknik *crawling* pada platform media sosial X. Dataset mentah yang diperoleh kemudian diproses melalui serangkaian prosedur *preprocessing* yang terdiri dari *cleansing*, *case folding*, normalisasi teks, *tokenizing*, *stopword removal*, hingga *stemming* untuk menjamin kualitas data yang akan diolah.



Gambar 1. Metode Penelitian

Setelah data dinyatakan bersih, dilakukan pelabelan menggunakan pendekatan berbasis leksikon serta eksplorasi data guna memahami distribusi kelas sentimen. Transformasi data teks ke dalam bentuk numerik dilakukan melalui metode pembobotan *Term Frequency Inverse Document Frequency* (TF-IDF), sebelum dataset dibagi secara proporsional menjadi data latih dan data uji. Tahap inti penelitian difokuskan pada pengembangan model klasifikasi dengan menerapkan tiga algoritma, yaitu *Decision Tree*, *Support Vector Machine* (SVM), dan *Random Forest*. Prosedur ini diakhiri dengan evaluasi kinerja model untuk mengukur tingkat presisi dan akurasi setiap algoritma dalam mengidentifikasi polaritas opini publik secara otomatis.

2.2 Pengumpulan D1ata

Penelitian ini memfokuskan observasi pada dinamika opini masyarakat di platform media sosial X terkait diskursus industri kelapa sawit dalam pusran kebijakan pemerintah yang berkaitan dengan figur Prabowo Subianto. Dataset yang diakuisisi terdiri dari kumpulan *tweet* yang merepresentasikan berbagai bentuk ekspresi publik, mulai dari dukungan terhadap program hilirisasi sawit hingga aspirasi terkait kesejahteraan petani. Proses perolehan data dilakukan melalui teknik *crawling* dengan mendayagunakan antarmuka pemrograman aplikasi (API) resmi X, di mana *auth token* digunakan sebagai instrumen autentikasi keamanan untuk menjamin legalitas akses data [11], [12]. Kriteria penyaringan data ditetapkan menggunakan kata kunci strategis seperti "Sawit" dan "Prabowo" dengan batasan waktu pengumpulan data secara longitudinal dari periode Januari hingga September 2025 guna menangkap spektrum sentimen publik secara komprehensif [13]. Seluruh data yang berhasil ditarik kemudian disimpan dalam format *Comma Separated Values* (CSV) untuk memfasilitasi integrasi data pada tahap pemrosesan selanjutnya. Struktur atribut data dipaparkan 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 terdahulu.
11	reply_count	Integer	Jumlah balasan terhadap tweet.
12	retweet_count	Integer	Jumlah retweet yang di terima 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 menyajikan rincian komprehensif mengenai karakteristik setiap atribut dalam dataset penelitian. Penjelasan tersebut mencakup identitas kolom, klasifikasi tipe data, tingkat validitas data, serta signifikansi fungsinya dalam mendukung proses analisis sentimen terhadap isu industri sawit.

2.3 Pre-Processing

Tahap *pre-processing* diimplementasikan untuk mentransformasi data tekstual mentah menjadi dataset yang siap olah. Prosedur pembersihan data dilakukan melalui beberapa tahapan sistematis untuk memastikan kualitas data yang akan diolah. Tahap pertama diawali dengan *cleaning data* untuk menghapus elemen non-esensial seperti *URL*, *mention*, *hashtag*, dan simbol khusus guna menyisakan teks murni. Selanjutnya, dilakukan *case folding* dengan mentransformasi seluruh huruf dalam *dataset* menjadi huruf kecil (*lowercase*) untuk menyeragamkan fitur kata. Proses kemudian dilanjutkan dengan normalisasi teks untuk memperbaiki kata tidak baku atau *slang*, seperti mengubah kata "*biodyzel*" menjadi "*biodiesel*" agar sesuai dengan standar bahasa Indonesia. Setelah teks seragam, tahap *tokenizing* dilakukan untuk memecah rangkaian kalimat menjadi satuan kata tunggal (*token*) agar dapat dianalisis secara individual. Untuk mempertajam analisis, dilakukan *stopword removal* guna mengeliminasi kata sambung umum yang tidak memiliki bobot sentimen signifikan, seperti "dan", "di", atau "yang". Rangkaian prosedur ini diakhiri dengan proses *stemming*, yaitu menghilangkan afiks untuk mengembalikan setiap kata ke bentuk dasarnya (*root word*), contohnya mengubah "perkebunan" menjadi "kebun", yang bertujuan untuk mereduksi dimensi data secara efektif [14], [15].

2.4 Pengolahan Data

Tahap pengolahan data dilakukan dengan tujuan untuk memberikan label sentimen, mengekstraksi fitur tekstual, dan menyiapkan struktur data agar dapat digunakan dalam proses komputasi model klasifikasi [11]. Adapun rangkaian tahapan yang dilakukan meliputi:

a. Pelabelan Data

Pada tahap ini, setiap teks ulasan mengenai industri sawit diberi label sentimen (positif atau negatif) menggunakan pendekatan *lexicon-based*. Proses ini dilakukan dengan mencocokkan kata-kata dalam tweet dengan kamus sentimen Bahasa Indonesia yang telah memiliki skor polaritas [18]. Nilai akumulasi polaritas dari setiap kata digunakan untuk menentukan label akhir pada setiap data opini terkait kebijakan sawit Prabowo Subianto [12].

b. Eksplorasi Data

Tahap ini bertujuan untuk memahami karakteristik dataset secara menyeluruh melalui analisis deskriptif, seperti distribusi jumlah data per kelas sentimen, panjang rata-rata teks, serta identifikasi kata kunci yang dominan muncul. Hasil eksplorasi digunakan untuk mendeteksi adanya ketidakseimbangan kelas (*imbalanced class*) serta memvisualisasikan kata yang paling sering muncul melalui *wordcloud* guna melihat tren opini publik terhadap isu sawit [13], [16].

c. Feature Extraction (TF-IDF)

Setelah data terlabeli, teks diubah menjadi representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Metode ini mengukur tingkat kepentingan suatu kata dalam dokumen dengan mempertimbangkan frekuensi kemunculannya dibandingkan dengan seluruh korpus dokumen dalam dataset [14], [19], [20].

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

Berdasarkan Persamaan (1) menjelaskan bahwa W_{ij} merupakan bobot suatu kata i pada dokumen j , dimana tf_{ij} menunjukkan frekuensi kemunculan kata tersebut dalam dokumen. Sementara itu, df_i menggambarkan jumlah dokumen yang mengandung kata i , dan N merupakan total keseluruhan dokumen dalam korpus penelitian industri sawit ini.

d. Split Dataset

Tahap terakhir dalam pengolahan data adalah membagi dataset menjadi dua bagian utama, yaitu data latih (*training set*) dan data uji (*testing set*) [14]. Pembagian ini dilakukan secara proporsional agar model dapat mempelajari pola dari sebagian besar data dan kemudian dievaluasi menggunakan data uji yang belum pernah diproses sebelumnya. Hal ini bertujuan untuk memastikan hasil prediksi model (SVM, *Random Forest*, dan *Decision Tree*) bersifat objektif dan memiliki generalisasi yang baik terhadap data baru [15].

2.5 Hasil dan Analisa

a. Modeling

Pada tahap ini, dataset ulasan industri sawit yang telah melewati proses ekstraksi fitur (TF-IDF) dan pembagian data (*split dataset*) digunakan sebagai basis pelatihan model. Penelitian ini mendayagunakan tiga algoritma pembelajaran mesin untuk dikomparasikan performanya, yaitu *Decision Tree* (DT), *Support Vector Machine* (SVM), dan *Random Forest* (RF) [19]. Pemilihan ketiga algoritma ini didasarkan pada efektivitasnya dalam menangani data teks berdimensi tinggi serta kemampuannya dalam memetakan polaritas sentimen publik terhadap isu kebijakan sawit secara otomatis.

b. Penerapan Algoritma *Decision Tree* (DT)

Decision Tree merupakan model klasifikasi yang membangun struktur hirarkis berbentuk pohon keputusan berdasarkan aturan fitur. Algoritma ini melakukan partisi data secara rekursif menggunakan pengukuran *Gini Impurity* untuk menentukan pemisahan (*split*) terbaik pada setiap node [10], [18]. Formulasi *Gini Impurity* didefinisikan sebagai:

$$Gini(D) = 1 - \sum_{i=1}^k p_i^2 \tag{2}$$

atau dengan Entropy:

$$Entropy(D) = - \sum_{i=1}^k p_i^2 \log_2 p_i \tag{3}$$

Dalam Persamaan (2), p_i merupakan probabilitas kelas ke- i dalam suatu node. Semakin kecil nilai Gini, semakin tinggi tingkat kemurnian data dalam node tersebut. Pada penelitian ini, DT mencatatkan akurasi sebesar 84%.

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

Algoritma SVM dioptimasi untuk memetakan opini publik terhadap kebijakan sawit dengan mencari *hyperplane* fungsional terbaik yang memisahkan kelas positif dan negatif dengan margin maksimal [11], [17]. Persamaan dasar SVM direpresentasikan sebagai:

$$f(x) = w^T x + b \tag{4}$$

Melalui Persamaan (3), w adalah vektor bobot dan x adalah fitur teks ulasan sawit. Penggunaan fungsi *kernel* pada SVM terbukti sangat efektif dalam menangani kompleksitas data tekstual sehingga berhasil mencapai performa tertinggi dengan akurasi sebesar 90%.

d. Penerapan Algoritma *Random Forest* (RF)

Random Forest diimplementasikan sebagai metode *ensemble learning* yang menggabungkan hasil dari banyak pohon keputusan (*decision trees*) yang dilatih pada subset data berbeda. Hasil akhir klasifikasi ditentukan melalui mekanisme *majority voting* guna meningkatkan stabilitas dan akurasi prediksi dibandingkan pohon tunggal [10], [18]. Formulasi *Random Forest* dituliskan sebagai:

$$y = mode(h_1(x), h_2(x), \dots, h_n(x)) \tag{5}$$

Dalam Persamaan (4), $h_i(x)$ adalah hasil prediksi dari tiap pohon, sementara \hat{y} merupakan label sentimen akhir (positif/negatif) berdasarkan suara terbanyak. Metode ini menghasilkan akurasi sebesar 85%

e. Evaluasi Model

Evaluasi model bertujuan untuk menilai sejauh mana algoritma mampu mengklasifikasikan sentimen masyarakat dengan benar melalui instrumen *confusion matrix* [16]

Tabel 2. Model Confusion Matrix

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

1. Accuracy

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{6}$$

Accuracy merupakan metrik yang mengukur tingkat ketepatan model dalam memprediksi label sentimen secara benar, baik untuk kategori positif maupun negatif. Dalam penelitian ini, nilai akurasi merepresentasikan kemampuan umum model dalam mengklasifikasikan seluruh opini publik terkait industri sawit secara presisi. Semakin tinggi nilai akurasi, semakin andal performa model dalam mengenali pola polaritas data secara keseluruhan [19].

2. Precison

$$Precision = \frac{TP}{TP+FP} \tag{7}$$



Precision menunjukkan rasio ketepatan antara data yang diprediksi positif dengan data yang sebenarnya positif. Dalam konteks isu industri sawit, nilai precision yang tinggi mengindikasikan bahwa model memiliki tingkat kesalahan yang rendah dalam mengklasifikasikan ulasan yang sebenarnya bersifat negatif namun terdeteksi sebagai positif. Hal ini krusial untuk menjamin bahwa dukungan publik yang terdeteksi benar-benar representatif [20].

3. Recall

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

Recall atau sensitivitas mengukur keberhasilan model dalam mengidentifikasi kembali seluruh data yang termasuk dalam kelas positif yang ada pada dataset. Nilai recall yang tinggi menunjukkan bahwa sebagian besar ulasan positif mengenai kebijakan sawit dan figur Prabowo Subianto berhasil ditemukan dan diklasifikasikan dengan benar oleh sistem tanpa banyak yang terlewatkan [20].

4. F1-Score

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{9}$$

F1-Score merupakan rata-rata harmonik yang mengintegrasikan komponen *precision* dan *recall* ke dalam satu nilai tunggal. Metrik ini memberikan gambaran yang lebih seimbang mengenai performa model, terutama jika terdapat ketidakseimbangan jumlah data antara sentimen positif dan negatif. Dalam penelitian ini, F1-Score merepresentasikan efektivitas model dalam menangani sentimen industri sawit secara adil dan stabil[20].

3. HASIL DAN PEMBAHASAN.

3.1 Pengumpulan Data

Data pada penelitian ini diperoleh dari hasil crawling tweet pengguna platform X yang membahas topik industri sawit dan Prabowo Subianto pada periode 1 Januari 2025 sampai 8 Maret 2026. Pengambilan data dilakukan secara otomatis dengan teknik web crawling menggunakan bantuan tweet-harvest berbasis Node.js serta bahasa pemrograman Python. Proses crawling difokuskan pada kata kunci yang relevan dengan topik penelitian untuk memperoleh data yang sesuai dengan tujuan analisis. Dari proses tersebut, berhasil dikumpulkan sebanyak 3.750 data tweet. Data yang terkumpul selanjutnya dikelola menggunakan library pandas dalam bentuk DataFrame agar memudahkan proses seleksi, pembersihan, dan transformasi data. Seluruh dataset kemudian disimpan dalam format CSV dan Excel. Secara keseluruhan, dataset memiliki 3.750 baris dan 16 kolom, dan selanjutnya digunakan dalam proses analisis sentimen, mulai dari tahap preprocessing, ekstraksi fitur TF-IDF, hingga klasifikasi machine learning menggunakan metode Decision Tree, Support Vector Machine, dan Random Forest.

3.2 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

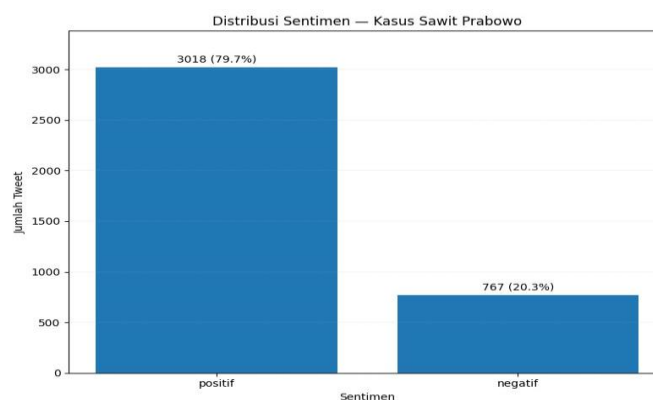
Alur Proses	Hasil
Teks Asli	Cintaku pada bagaikan Prabowo dengan sawit menyebar luas dimanaa mana hingga ke pelosok negeri dimna ada lahan kosong itu pasti ada sawit oh sayangku sawit yang kucinta
Cleaning Data	Cintaku pada bagaikan Prabowo dengan sawit menyebar luas dimanaa mana hingga ke pelosok negeri dimna ada lahan kosong itu pasti ada sawit oh sayangku sawit yang kucinta
Case Folding	cintaku pada bagaikan prabowo dengan sawit menyebar luas dimanaa mana hingga ke pelosok negeri dimna ada lahan kosong itu pasti ada sawit oh sayangku sawit yang kucinta
Normalisasi Teks	cintaku pada bagaikan prabowo dengan sawit menyebar luas dimanaa mana hingga ke pelosok negeri dimna ada lahan kosong itu pasti ada sawit oh sayangku sawit yang kucinta
Tokenizing	cintaku', 'pada', 'bagaikan', 'prabowo', 'dengan', 'sawit', 'menyebar', 'luas', 'dimanaa', 'mana', 'hingga', 'ke', 'pelosok', 'negeri', 'dimna', 'ada', 'lahan', 'kosong', 'itu', 'pasti', 'ada', 'sawit', 'oh', 'sayangku', 'sawit', 'yang', 'kucinta'
Stopword Removal	'cintaku', 'prabowo', 'sawit', 'menyebar', 'luas', 'dimanaa', 'pelosok', 'negeri', 'dimna', 'lahan', 'kosong', 'sawit', 'oh', 'sayangku', 'sawit', 'kucinta'
Stemming	'cinta', 'prabowo', 'sawit', 'sebar', 'luas', 'dimanaa', 'pelosok', 'negeri', 'dimna', 'lahan', 'kosong', 'sawit', 'oh', 'sayang', 'sawit', 'cinta'
Stemming String	cinta prabowo sawit sebar luas dimanaa pelosok negeri dimna lahan kosong sawit oh sayang sawit cinta

3.3 Pengolahan Data

Tahap pengolahan data diawali dengan pelabelan otomatis menggunakan pendekatan *Lexicon-Based* untuk menentukan kategori sentimen positif, negatif, atau netral, yang kemudian dilanjutkan dengan eksplorasi data untuk memahami karakteristik serta tren kata dalam dataset. Selanjutnya, dilakukan ekstraksi fitur menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* untuk mentransformasi teks menjadi representasi numerik berdasarkan bobot kepentingan kata, dan diakhiri dengan pembagian dataset (*Split Dataset*) ke dalam porsi data latih dan data uji guna memastikan model klasifikasi dapat dikembangkan dan dievaluasi secara akurat.

3.3.1 Pelabelan Data

Setelah melalui tahap preprocessing, data tweet selanjutnya diberi label sentimen menggunakan pendekatan lexiconbased. Skor sentimen diperoleh dari hasil penjumlahan bobot kata berdasarkan kamus sentimen yang digunakan dalam penelitian. Tweet diklasifikasikan ke dalam sentimen positif apabila memiliki skor ≥ 0 , sedangkan tweet dengan skor < 0 diklasifikasikan sebagai sentimen negatif. Berdasarkan hasil pelabelan, diperoleh 3.018 tweet (79,7%) berkategori positif dan 767 tweet (20,3%) berkategori negatif dari total 3.785 tweet. Hasil ini menunjukkan bahwa opini publik pada data yang dikumpulkan cenderung didominasi oleh sentimen positif pada Gambar 2.



Gambar 2. Hasil Pelabelan Data

Berdasarkan Gambar 2, tahapan dimulai dengan pengumpulan data, dilanjutkan dengan *pre-processing*, pelabelan, hingga ekstraksi fitur TF-IDF. Setelah model dibangun menggunakan algoritma *Machine Learning*, tahap terakhir adalah pengujian performa untuk mengukur tingkat *Accuracy* (Akurasi). Akurasi ini digunakan untuk melihat sejauh mana model mampu mengklasifikasikan sentimen masyarakat terhadap industri sawit dan Prabowo Subianto secara tepat dengan membandingkan hasil prediksi model terhadap label asli pada data uji.

Tabel 4. Hasil Pelabelan Data

Teks	Skor Sentimen	Label Sentimen
netizen dikit komen sawitceo sawitga nyadar apa mrk ngomong gtu pdhl keseharian pengguna sawit dr minyak goreng margarin sabun shampo lipstick roti lotion deterjen kn itu dr sawit prabowo	-5	Negatif
Pemerintah harus membangun beberapa pabrik CPO di beberapa titik+1 perkebunan sawit agar harga tidak dipermainkan oleh pengusaha besar yang punya pabrik. Selama ini petani sawit sering menjadi korban permainan harga oleh pemilik pabrik		Positif
Sawit Jadi Tameng Energi Indonesia di Tengah Konflik Timur Tengah. Ini+5 Buktinya. Jakarta Pernyataan Presiden Prabowo Subianto yang menyebut kelapa sawit sebagai tanaman ajaib sempat menuai sindiran di media sosial, namun di tengah eskalasi perang di Timur Tengah sawit		Positif
Betul bro tarif dua arah bikin produk AS dan Indo kompetitif sama di pasar+11 masing-masing. Tapi ini peluang besar UMKM Indo, sawit, kopi, rempah, elektronik dan lainnya bisa ekspor bebas ke pasar AS tanpa bea masuk. Di dalam negeri kompetisi sehat mendorong inovasi dan kualitas naik		Positif
klo pohon sawit kaya liat prabowo	-4	Negatif

Tabel 4 menyajikan beberapa contoh data tweet yang telah melalui proses pelabelan sentimen pada topik industri sawit dan Prabowo Subianto. Setiap tweet diberikan skor sentimen berdasarkan akumulasi bobot kata yang terdapat dalam kamus sentimen menggunakan metode lexicon-based. Berdasarkan skor yang diperoleh, tweet kemudian diklasifikasikan ke dalam dua kategori sentimen, yaitu positif dan negatif.

3.3.2 Eksplorasi Data

Tahap eksplorasi data dilakukan untuk memahami karakteristik awal dataset sebelum masuk ke proses pemodelan. Salah satu bentuk eksplorasi yang digunakan dalam penelitian ini adalah visualisasi WordCloud, yang berfungsi untuk menampilkan kata-kata yang paling dominan pada tweet terkait isu industri sawit dan Prabowo Subianto. Melalui visualisasi ini, dapat diketahui kecenderungan kata yang sering muncul pada masing-masing kelas sentimen. WordCloud pada tweet berlabel positif dan negatif digunakan untuk memberikan gambaran awal mengenai pola bahasa yang muncul dalam data. Hasil visualisasi WordCloud sentimen positif ditampilkan pada Gambar 3, sedangkan WordCloud sentimen negatif ditampilkan pada Gambar 4.



Gambar 3. Wordcloud Sentimen Positif



Gambar 4. Wordcloud Sentimen Negatif

Gambar 3 memperlihatkan visualisasi WordCloud dari tweet dengan sentimen positif pada topik industri sawit dan Prabowo Subianto. Kata-kata yang tampak dominan antara lain “prabowo”, “sawit”, “program”, “indonesia”, “presiden”, dan “rakyat”. Dominasi kata-kata tersebut menunjukkan bahwa sentimen positif dalam dataset cenderung berkaitan dengan dukungan terhadap program pemerintah, kebijakan, serta persepsi yang menghubungkan Prabowo dengan peran sawit dalam pembangunan dan perekonomian nasional. Sedangkan Gambar 4 menampilkan visualisasi WordCloud dari tweet dengan sentimen negatif pada topik yang sama. Kata-kata dominan yang muncul antara lain “prabowo”, “sawit”, “mbak”, “amp”, “minyak”, “indonesia”, dan “negara”. Hal ini menunjukkan bahwa tweet negatif lebih banyak mengandung kritik, sindiran, maupun opini kontra terhadap isu sawit dan tokoh yang dibahas, sehingga mencerminkan adanya persepsi negatif pada sebagian percakapan publik di media sosial.

3.3.3 Feature Extraction

Berdasarkan hasil pembobotan TF-IDF pada Tabel 5, kata-kata dengan bobot tertinggi antara lain “prabowo”, “program”, “sawit”, “indonesia”, dan “bijak”. Kemunculan kata-kata tersebut menunjukkan bahwa pembahasan dalam dataset didominasi oleh isu yang berkaitan dengan Prabowo Subianto, program pemerintah, serta industri sawit dalam konteks nasional. Sementara itu, kata-kata dengan bobot terendah cenderung merupakan kata yang jarang muncul dalam dokumen, sehingga kontribusinya terhadap pembentukan fitur juga relatif kecil. Hasil ini menunjukkan bahwa metode TF-IDF mampu menyortir kata-kata yang paling representatif dalam dataset dan layak digunakan sebagai fitur pada tahap klasifikasi sentimen.

Tabel 5. Hasil TF-IDF

No	Term	TF-IDF	TF
0	prabowo	162.167951	1743
1	program	139.008641	1087
2	sawit	121.431463	620
3	indonesia	106.829684	794
4	bijak	104.934712	777
...
7126	Jatiwaringin	0.134988	1
7127	wisma	0.134988	1
7128	tebet	0.134988	1
7129	kcp	0.134988	1
7130	sudirman	0.134988	1

3.3.4 Split Data

Pada tahap ini, dataset dibagi menjadi dua bagian utama, yaitu data latih (training data) dan data uji (testing data), untuk mendukung proses pembangunan serta evaluasi model klasifikasi. Pembagian dilakukan berdasarkan label sentimen agar distribusi kelas pada data latih dan data uji tetap terjaga. Hasil split dataset menunjukkan bahwa data latih terdiri dari 2.414 data positif dan 614 data negatif, sedangkan data uji terdiri dari 604 data positif dan 153 data negatif. Ringkasan jumlah data pada masing-masing subset ditampilkan pada Tabel 6.

Tabel 6. Hasil Split Data

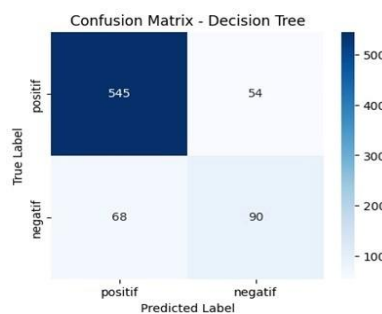
Dataset	Label	Jumlah
Latih	Positif	2414
Latih	Negatif	614
Uji	Positif	604
Uji	Negatif	153

3.4 Hasil dan Analisa

Pada tahap ini, dilakukan analisis terhadap hasil klasifikasi sentimen pada dataset tweet mengenai industri sawit dan Prabowo Subianto yang diperoleh melalui proses crawling dari platform X. Untuk mengukur performa model klasifikasi yang digunakan, penelitian ini memanfaatkan Confusion Matrix (CM) yang terdiri atas True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). Melalui Confusion Matrix, dapat diketahui tingkat ketepatan model dalam membedakan tweet yang termasuk ke dalam sentimen positif maupun negatif.

3.4.1 Decision Tree

Evaluasi performa model dilakukan menggunakan Confusion Matrix untuk melihat tingkat ketepatan prediksi pada tiap kelas sentimen. Pada penelitian ini, model Decision Tree digunakan untuk mengklasifikasikan sentimen tweet ke dalam kategori positif dan negatif. Hasil prediksi model Decision Tree terhadap data uji dapat dilihat pada Gambar 5.



Gambar 5. Confusion Matrix - Decision Tree

Gambar 5 menampilkan Confusion Matrix model Decision Tree dengan hasil TP = 545, TN = 90, FP = 68, dan FN = 54, yang menunjukkan bahwa model lebih baik dalam mengenali sentimen positif dibandingkan sentimen negatif.

```

*** Decision Tree ***
Akurasi: 0.8388375165125496
precision    recall  f1-score   support

negatif      0.62    0.57    0.60     158
positif      0.89    0.91    0.90     599

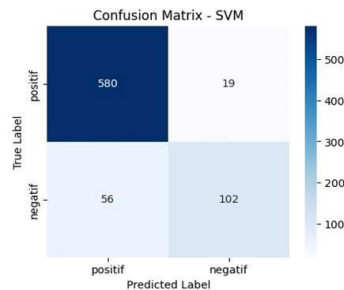
accuracy          0.84     757
macro avg         0.76    0.74    0.75     757
weighted avg     0.83    0.84    0.84     757
    
```

Gambar 6. Evaluasi Model Decision Tree

Berdasarkan Gambar 6, model Decision Tree memperoleh akurasi sebesar 0,84 dengan nilai precision sebesar 0,62 untuk kelas negatif dan 0,89 untuk kelas positif. Nilai recall menunjukkan bahwa model lebih baik dalam mengenali sentimen positif, yaitu sebesar 0,91, dibandingkan sentimen negatif sebesar 0,57. Nilai f1-score yang diperoleh adalah 0,60 untuk kelas negatif dan 0,90 untuk kelas positif. Secara keseluruhan, performa model Decision Tree tergolong cukup baik, namun masih menunjukkan ketidakseimbangan dalam mendeteksi kelas negatif.

3.3.2 Support Vector Machine

Evaluasi performa model dilakukan menggunakan Confusion Matrix untuk melihat tingkat ketepatan prediksi pada tiap kelas sentimen. Pada penelitian ini, model Support Vector Machine (SVM) digunakan untuk mengklasifikasikan sentimen tweet ke dalam kategori positif dan negatif. Hasil prediksi model SVM terhadap data uji dapat dilihat pada Gambar 7.



Gambar 7. Confusion Matrix - Support Vector Machine

Gambar 7 menampilkan Confusion Matrix model SVM dengan hasil TP = 580, TN = 102, FP = 56, dan FN = 19, yang menunjukkan bahwa model lebih baik dalam mengenali sentimen positif dibandingkan sentimen negatif.

```

=== SVM ===
Akurasi: 0.9009247027741083
precision  recall  f1-score  support
negatif    0.84    0.65    0.73    158
positif    0.91    0.97    0.94    599

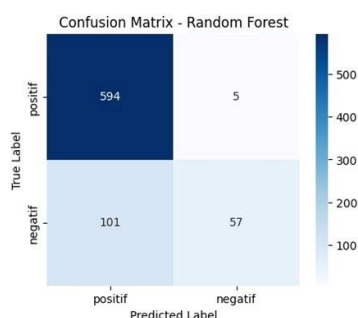
accuracy    0.90    757
macro avg   0.88    0.81    0.84    757
weighted avg 0.90    0.90    0.90    757
    
```

Gambar 8. Evaluasi Model Support Vector Machine

Berdasarkan Gambar 8, model SVM memperoleh akurasi sebesar 0,90 dengan nilai precision sebesar 0,84 untuk kelas negatif dan 0,91 untuk kelas positif. Nilai recall menunjukkan bahwa model lebih baik dalam mengenali sentimen positif, yaitu sebesar 0,97, dibandingkan sentimen negatif sebesar 0,65. Nilai f1-score yang diperoleh adalah 0,73 untuk kelas negatif dan 0,94 untuk kelas positif. Secara keseluruhan, performa model SVM tergolong sangat baik dan menunjukkan hasil yang lebih stabil dalam mengklasifikasikan sentimen positif maupun negatif dibandingkan model sebelumnya.

3.4.3 Random Forest

Evaluasi performa model dilakukan menggunakan Confusion Matrix untuk melihat tingkat ketepatan prediksi pada tiap kelas sentimen. Pada penelitian ini, model Random Forest digunakan untuk mengklasifikasikan sentimen tweet ke dalam kategori positif dan negatif. Hasil prediksi model Random Forest terhadap data uji dapat dilihat pada Gambar 9.



Gambar 9. Confusion Matrix - Random Forest

Gambar 9 menampilkan Confusion Matrix model Random Forest dengan hasil TP = 594, TN = 57, FP = 101, dan FN = 5, yang menunjukkan bahwa model sangat baik dalam mengenali sentimen positif, tetapi masih kurang optimal dalam mendeteksi sentimen negatif.

```

=== Random Forest ===
Akurasi: 0.8599735799207398
precision  recall  f1-score  support
negatif    0.92    0.36    0.52    158
positif    0.85    0.99    0.92    599

accuracy    0.86    757
macro avg   0.89    0.68    0.72    757
weighted avg 0.87    0.86    0.83    757
    
```

Gambar 10. Evaluasi Model Random Forest

Berdasarkan Gambar 10, model Random Forest memperoleh akurasi sebesar 0,86 dengan nilai precision sebesar 0,92 untuk kelas negatif dan 0,85 untuk kelas positif. Nilai recall menunjukkan bahwa model jauh lebih baik dalam mengenali sentimen positif, yaitu sebesar 0,99, dibandingkan sentimen negatif sebesar 0,36. Nilai f1-score yang diperoleh adalah 0,52 untuk kelas negatif dan 0,92 untuk kelas positif. Secara keseluruhan, performa model Random Forest tergolong baik dari sisi akurasi, namun masih menunjukkan ketidakseimbangan yang cukup besar dalam mendeteksi kelas negative

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
Decision Tree	545	90	68	54
Support Vector Machine	580	102	56	19
Random Forest	594	57	101	5

Hal ini ditunjukkan oleh jumlah prediksi benar yang paling tinggi, yaitu True Positive sebesar 580 dan True Negative sebesar 102. Sementara itu, Random Forest memiliki True Positive tertinggi, namun juga menghasilkan False Positive yang paling tinggi sehingga performanya cenderung kurang seimbang. Decision Tree menunjukkan performa yang lebih rendah dibandingkan kedua model lainnya. Dengan demikian, berdasarkan analisis confusion matrix, SVM dapat dinilai sebagai model yang paling optimal pada penelitian ini.

Tabel 8. Hasil Perbandingan Model

Model	Class	Accuracy	Precision	Recall	F1-score
Decision Tree	Negative	0.84	0.62	0.57	0.60
	Positive		0.89	0.91	0.90
Support Vector Machine	Negative	0.90	0.84	0.65	0.73
	Positive		0.91	0.97	0.94
Random Forest	Negative	0.86	0.92	0.36	0.52
	Positive		0.85	0.99	0.92

Berdasarkan hasil perbandingan model pada Tabel 8, algoritma Support Vector Machine (SVM) menunjukkan kinerja terbaik dengan akurasi tertinggi sebesar 0.90 pada kedua kelas. Model ini juga memperoleh nilai precision, recall, dan F1-score yang tinggi, khususnya pada kelas positif dengan nilai masing-masing sebesar 0.91, 0.97, dan 0.94. Hal ini menunjukkan bahwa model SVM mampu mengklasifikasikan data secara lebih baik dan konsisten dibandingkan model lainnya. Sementara itu, Random Forest menempati posisi kedua dengan akurasi sebesar 0.86, sedangkan Decision Tree memiliki performa yang lebih rendah meskipun masih menunjukkan hasil klasifikasi yang cukup baik. Dengan demikian, hasil ini mengindikasikan bahwa model SVM merupakan model yang paling optimal dalam mengklasifikasikan sentimen pada dataset penelitian ini.

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa analisis sentimen terhadap opini publik di platform X mengenai industri kelapa sawit dan Prabowo Subianto dapat dilakukan secara efektif menggunakan pendekatan machine learning berbasis fitur TF-IDF. Tahapan preprocessing yang meliputi cleaning, case folding, normalisasi teks, tokenizing, stopword removal, dan stemming terbukti membantu menyiapkan data teks yang lebih terstruktur untuk proses klasifikasi. Dari tiga algoritma yang dibandingkan, yaitu Decision Tree, Random Forest, dan Support Vector Machine, model SVM menunjukkan performa terbaik dengan akurasi sebesar 90%, lebih tinggi dibandingkan Random Forest sebesar 86% dan Decision Tree sebesar 84%. Selain memiliki akurasi tertinggi, SVM juga memperlihatkan performa yang lebih stabil pada metrik precision, recall, dan F1-score, sehingga dinilai paling optimal dalam mengklasifikasikan sentimen positif dan negatif pada dataset penelitian ini. Hasil pelabelan data juga menunjukkan bahwa percakapan publik cenderung didominasi oleh sentimen positif, yang mengindikasikan adanya respons masyarakat yang relatif baik terhadap isu industri sawit dalam konteks kebijakan yang dikaitkan dengan Prabowo Subianto. Meskipun demikian, penelitian ini masih memiliki keterbatasan, terutama pada distribusi kelas yang tidak seimbang dan penggunaan pelabelan otomatis berbasis leksikon yang belum sepenuhnya mampu menangkap ironi, sarkasme, atau konteks bahasa informal media sosial. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan teknik penyeimbangan data, memperluas jumlah dataset, serta membandingkan model machine learning dengan pendekatan deep learning agar hasil klasifikasi menjadi lebih akurat dan robust.

REFERENCES

- [1] M. Ma'rufudin and A. Yudhistira, "Analisis Sentimen Petani Milenial Pada Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM)," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 3, pp. 845–857, 2025, doi: 10.52436/1.jpti.717.



- [2] A. Ma'rif, A. E. Pajri, and P. Liana, "Sentiment Analysis of President Prabowo's Performance on Twitter (X) with a Comparative Study of SVM, XGBoost, and AdaBoost," *Journal of Applied Informatics and Computing (JAIC)*, vol. 10, no. 1, pp. 684–697, Feb. 2026, doi: 10.30871/jaic.v10i1.12138.
- [3] T. Wahyudi, Rudiman, and N. A. Verdikha, "Klasifikasi sentimen X-Twitter perihal pemindahan ibu kota Indonesia menggunakan ekstraksi fitur TF-IDF dan metode Support Vector Machine (SVM)," *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 18, no. 2, pp. 185–199, 2024, doi: 10.47111/jti.v18i2.15015.
- [4] N. Nur, A. Aryanti, and O. Suria, "Analisis sentimen terhadap keputusan hubungan kerja di Indonesia: Komparasi IndoBERT dengan SVM, Random Forest, dan Decision Tree dengan optimasi TF-IDF," *Building of Informatics, Technology and Science (BITS)*, vol. 10, no. 2, pp. 1158–1176, 2025, doi: 10.47065/bits.v10i2.999.
- [5] V. B. Lestari and C. A. Hutagalung, "Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes," *J. Comput. Sci. Appl.*, vol. 8, no. 1, pp. 22–2025, 2025, [Online]. Available: <https://doi.org/10.21009/j->
- [6] M. N. R. N. Suryono, A. A. Arifiyanti, and D. S. Y. Kartika, "Aspect-based sentiment analysis pada ulasan aplikasi Access by KAI menggunakan metode TF-IDF dan algoritma Support Vector Machine," *Jurnal INSTEK: Informatika Sains dan Teknologi*, vol. 10, no. 2, pp. 414–423, Oct. 2025, doi: 10.24252/instek.v10i2.57155.
- [7] R. D. Cahyani and P. T. Prasetyaningrum, "Sentiment Analysis of User Reviews for AI Applications : Evaluating SVM , Logistic Regression , and Random Forest," vol. 8, no. 1, pp. 1–27, 2026, doi: 10.63158/journalisi.v8i1.1366.
- [8] U. A. Pringsewu, S. Pahrullah, D. Feriyanto, T. H. Andika, and N. Aminudin, "Perbandingan Dan Efektivitas Kinerja Algoritma Svm Dan Bi-Lstm Dengan Tf-Idf Dan Smote Untuk Analisis Sentimen Kelestarian Hewan Liar," vol. 7, no. 1, pp. 9–14.
- [9] J. E. Br Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF," *J. Manaj. Inform.*, vol. 14, no. 1, pp. 42–53, 2024, doi: 10.34010/jamika.v14i1.11946.
- [10] Indriani and A. Davy Wiranata, "Perbandingan Tingkat Akurasi Algoritma Svm, Decision Tree, Dan Random Forest Dalam Analisis Sentimen Tanggapan Pengguna Aplikasi Gopay," *J. Tek. Inform.*, vol. 5, no. 3, pp. 777–787, 2024, [Online]. Available: <https://doi.org/10.52436/1.jutif.2024.5.3.1885>
- [11] B. M. Obidje and M. A. I. Pakereng, "Analisis Sentimen Pemilihan Presiden dan Wakil Presiden Tahun 2024 Di Twitter Menggunakan Metode Klasifikasi Naive Bayes," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 1, pp. 424–433, 2025, doi: 10.29100/jupi.v10i1.5836.
- [12] M. R. Saputra and P. Parjito, "Analisis Sentimen Twitter Terhadap Konflik Di Papua Menggunakan Perbandingan Naive Bayes Dan Svm," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 2, pp. 1197–1208, 2025, doi: 10.29100/jupi.v10i2.6180.
- [13] A. B. Aditya, S. Samsudin, W. P. Rizki, M. Mahendra, and A. Setiawan, "Analisis Sentimen Ulasan Aplikasi Gojek Menggunakan Support Vector Machine Dan Random Forest," *J. Inform. Terpadu*, vol. 11, no. 2, pp. 134–143, 2025, doi: 10.54914/jit.v11i2.1884.
- [14] V. A. Sulistiani and M. Hamka, "Analisis Sentimen Pengguna Media Sosial Terhadap Identitas Kependudukan Digital Menggunakan Metode Support Vector Machine (SVM)," *J. Inf. Syst. Res.*, vol. 5, no. 4, pp. 1323–1332, 2024, doi: 10.47065/josh.v5i4.5614.
- [15] F. Diannita Sari, W. Widayat, and R. Artikel, "Analisis sentimen masyarakat terhadap PON 2024 Aceh dan Sumatra Utara di Twitter menggunakan algoritma SVM dan KNN," *AITI J. Teknol. Inf.*, vol. 22, no. 2, pp. 102–116, 2025.
- [16] R. F. Apriyani, D. A. Megawaty, S. Informasi, U. T. Indonesia, B. Lampung, and K. Merdeka, "Komparasi Performa Klasifikasi Sentimen Masyarakat Terhadap Kurikulum Merdeka di Sekolah Menggunakan SVM dan KNN," vol. 6, no. 4, pp. 2795–2806, 2025, doi: 10.47065/bits.v6i4.6877.
- [17] E. T. Pardede, K. D. Tania, and M. Afrina, "Knowledge Discovery: Analisis Sentimen dan Emosi WhatsApp Business dengan Machine learning dan Deep Learning," *J. Nas. Teknol. dan Sist. Inf.*, vol. 11, no. 3, pp. 310–318, 2026, doi: 10.25077.
- [18] A. Mustika Rani and N. Hendrastuty, "Perbandingan Algoritma NBC Dan SVM Untuk Melakukan Analisis Sentimen Terhadap PP NO.82 Tahun 2021," *Technol. Sci.*, vol. 6, no. 4, pp. 2139–2151, 2025, doi: 10.47065/bits.v6i4.6496.
- [19] N. C. Ramadani, I. Tahyudin, and A. Shouni Barkah, "Perbandingan Algoritma Support Vector Machine, Decision Tree, dan Logistic Regression Pada Analisis Sentimen Ulasan Aplikasi Netflix," *J. Nas. Teknol. dan Sist. Inf.*, vol. 10, no. 2, pp. 110–117, 2024, doi: 10.25077/teknosi.v10i2.2024.110-117.
- [20] P. Publik, S. Twitter, and J. Juliantono, "Persepsi Publik Terhadap Kepemimpinan Firli Bahuri di KPK: Pendekatan Sentimen Twitter dengan Naïve Bayes dan SVM," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1272–1285, 2025. [Online]. Available: <https://doi.org/10.29100/jupi.v10i2.6181>.