

# Analisis Sentimen Publik Konflik Agraria PT TPL dengan Masyarakat Adat Toba Simalungun Menggunakan Metode SVM

## *Public Sentiment Analysis of the Agrarian Conflict between PT TPL and the Toba Simalungun Indigenous Community Using the SVM Method*

Dian Yusri Andira<sup>1</sup>, Deswita Maharani Harahap<sup>2</sup>, Vibiola Br Damanik<sup>3</sup>, Indah Frian Sari<sup>4</sup>, Victor Asido Elyakim P<sup>5</sup>

<sup>1,2,3,4,5</sup>STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara

### Article Info

#### *Genesis Artikel:*

Diterima, 23 Juli 2025  
Direvisi, 07 Agustus 2025  
Disetujui, 15 September 2025

#### *Kata Kunci:*

Analisis Sentimen  
Agrarian  
Tiktok  
Support Vector Machine  
Tf-Idf

### ABSTRAK

Konflik agraria antara PT Toba Pulp Lestari dan masyarakat adat Toba Simalungun telah memunculkan beragam opini publik di media sosial. Penelitian ini bertujuan menganalisis sentimen publik terkait konflik tersebut menggunakan metode Support Vector Machine (SVM) berbasis data komentar TikTok. Sebanyak 1.751 komentar dikumpulkan melalui API TikTok dan diproses melalui pembersihan, normalisasi, penghapusan kata tidak bermakna, serta stemming. Pelabelan sentimen dilakukan secara otomatis dengan pendekatan leksikal, kemudian dilakukan pembobotan fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Model SVM digunakan untuk mengklasifikasikan sentimen publik ke dalam dua kategori utama, yaitu positif dan negatif. Hasil pengujian menunjukkan bahwa model SVM mampu mencapai akurasi sebesar 80%, dengan performa yang sangat baik dalam mendekripsi sentimen negatif. Analisis tambahan melalui visualisasi wordcloud menunjukkan kata-kata dominan pada masing-masing kategori sentimen yang memperkuat hasil klasifikasi model. Temuan penelitian ini memberikan gambaran objektif mengenai pola opini publik di media sosial, sekaligus menunjukkan potensi penerapan metode analisis sentimen berbasis pembelajaran mesin untuk memahami persepsi masyarakat terhadap isu sosial lainnya di masa mendatang.

### ABSTRACT

*The agrarian conflict between PT Toba Pulp Lestari and the Toba Simalungun indigenous community has generated diverse public opinions on social media. This study aims to analyze public sentiment regarding the conflict using the Support Vector Machine (SVM) method based on TikTok comment data. A total of 1,751 comments were collected via the TikTok API and processed through cleaning, normalization, stopword removal, and stemming. Sentiment labeling was performed automatically with a lexical-based approach, followed by feature weighting using Term Frequency-Inverse Document Frequency (TF-IDF). The SVM model was used to classify public sentiment into two main categories, namely positive and negative. The results of the testing showed that the SVM model was able to achieve an accuracy of 80%, with excellent performance in detecting negative sentiment. Additional analysis through wordcloud visualization shows the dominant words in each sentiment category, which reinforces the model's classification results. The findings of this study provide an objective picture of public opinion patterns on social media, while also demonstrating the potential application of machine learning-based sentiment analysis methods to understand public perceptions of other social issues in the future.*

*This is an open access article under the [CC BY-SA](#) license.*



#### *Corresponding Author:*

Dian Yusri Andira  
Program Studi Sistem Informasi, STIKOM Tunas Bangsa  
Jl. Jend. Sudirman, Blok A No. 1,2&3, Siantar Barat, Pematangsiantar, Indonesia.  
Email: dianyusriandira296@gmail.com

## 1. PENDAHULUAN

Konflik agraria di Indonesia hingga saat ini masih menjadi permasalahan struktural yang berdampak luas pada aspek sosial, ekonomi, budaya, dan ekologi, terutama di wilayah-wilayah yang dihuni masyarakat adat. Tanah, bagi masyarakat adat Toba Simalungun, tidak hanya berfungsi sebagai sumber penghidupan dan aset ekonomi, tetapi juga menjadi inti dari identitas budaya dan spiritual yang diwariskan lintas generasi [1]. Ketimpangan akses dan distribusi lahan, terutama akibat tumpang tindih klaim antara komunitas adat dan perusahaan besar yang memperoleh konsesi dari pemerintah, telah memicu banyak konflik yang berlarut-larut di berbagai daerah. Salah satu kasus yang paling mencolok adalah konflik antara masyarakat adat Pandumaan-Sipituhuta dengan PT Toba Pulp Lestari (TPL) di Kabupaten Humbang Hasundutan, Sumatera Utara. Konflik ini timbul akibat tumpang tindih kepemilikan lahan, di mana kawasan hutan adat yang secara turun-temurun dikelola sebagai hutan kemenyan ('tombak haminjon') oleh masyarakat adat beririsan langsung dengan area konsesi yang diberikan kepada perusahaan oleh negara [1]. Kebijakan ini menyebabkan terjadinya perlawanan sosial berupa protes, aksi blokade, hingga kriminalisasi masyarakat adat yang berupaya mempertahankan hak-hak leluhur mereka. Hutan kemenyan menjadi simbol ekonomi, budaya, dan spiritual masyarakat adat, sekaligus representasi identitas kolektif serta sumber perlawanan terhadap praktik-praktik marginalisasi yang mengatasnamakan pembangunan dan investasi [2]

Dalam konteks era digital saat ini, dinamika opini dan sentimen publik terhadap konflik agraria mengalami transformasi yang signifikan. Informasi, narasi, serta opini masyarakat tentang isu agraria kini banyak ditemukan di ruang digital, khususnya media sosial, sehingga membentuk diskursus baru dan memperluas pengaruhnya pada pengambilan kebijakan publik. Berbagai penelitian membuktikan bahwa analisis sentimen berbasis machine learning, seperti Support Vector Machine (SVM), terbukti efektif untuk memetakan opini dan persepsi masyarakat dalam isu-isu strategis, mulai dari kebijakan pemerintah, pertambangan, hingga pelayanan publik [3], [4], [5]. Penggunaan SVM dalam klasifikasi sentimen menawarkan keunggulan dalam hal akurasi, kecepatan, serta kemampuan menangani data yang besar dan tidak terstruktur.

Meski demikian, terdapat kesenjangan penelitian (research gap) yang cukup nyata dalam literatur. Kajian-kajian sebelumnya umumnya menyoroti aspek kronologis, pemetaan aktor, dan analisis struktural konflik agraria [1], sementara penelitian yang secara sistematis mengintegrasikan analisis sentimen publik berbasis media sosial dengan pendekatan machine learning (khususnya SVM) pada kasus-kasus konflik agraria lokal masih sangat terbatas. Selain itu, penerapan SVM dalam menganalisis opini masyarakat terhadap konflik agraria antara PT TPL dan masyarakat adat Toba Simalungun belum banyak dijumpai di literatur Indonesia. Padahal, analisis sentimen publik semacam ini dapat menjadi sumber data baru yang penting untuk advokasi, evaluasi kebijakan, dan penguatan suara masyarakat terdampak.

Dengan memperhatikan gap tersebut, penelitian ini mengusulkan pendekatan analisis sentimen publik berbasis Support Vector Machine terhadap konflik agraria antara PT Toba Pulp Lestari dan masyarakat adat Toba Simalungun. Penelitian ini diharapkan mampu memberikan gambaran yang objektif terkait persepsi masyarakat, memperkaya literatur nasional mengenai integrasi machine learning dalam studi agraria, serta memberikan kontribusi nyata bagi proses penyusunan kebijakan dan penyelesaian konflik agraria yang berkeadilan.

## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Pada penelitian ini, data dikumpulkan melalui pemanfaatan API resmi TikTok yang memungkinkan pengambilan komentar secara terstruktur dan sistematis dari platform tersebut. Proses pengambilan data difokuskan pada komentar-komentar yang berkaitan dengan isu konflik agraria antara PT Toba Pulp Lestari dan masyarakat adat Toba Simalungun. Kata kunci pencarian ditetapkan secara khusus untuk menangkap percakapan publik yang relevan, baik yang menyebutkan konflik secara langsung maupun secara tersirat. Pengumpulan data dilakukan pada saat intensitas diskusi mengenai konflik ini sedang meningkat di TikTok, sehingga respons yang terekam dapat merepresentasikan opini dan sentimen masyarakat secara aktual. Dari proses ini, diperoleh sebanyak 1.751 komentar yang selanjutnya dijadikan korpus utama untuk proses analisis sentimen menggunakan metode Support Vector Machine.

### 2.2 Text Preprocessing

Pengolahan data dimulai dengan menata teks mentah hasil pengambilan komentar dari TikTok, yang dilakukan pada lingkungan Google Colab. Dalam proses ini, digunakan berbagai pustaka Python seperti re untuk membersihkan karakter-karakter yang tidak diperlukan, nltk untuk menghilangkan kata-kata umum yang kurang bermakna (stopwords), serta Sastrawi untuk melakukan stemming terhadap kata-kata dalam Bahasa Indonesia. Langkah awal pembersihan mencakup penghapusan tautan, emotikon, mention akun, tagar, angka, serta simbol-simbol lain yang tidak relevan terhadap analisis sentimen. Selain itu, seluruh komentar dikonversi ke huruf kecil (lowercase) untuk menjaga konsistensi data.

Pada tahap stopword removal, daftar stopwords Bahasa Indonesia diperkaya dengan berbagai istilah informal yang sering muncul di media sosial, misalnya "loh", "dong", atau "nih", sehingga proses pembersihan dapat lebih optimal. Selanjutnya, proses stemming dilakukan untuk mengembalikan setiap kata ke bentuk dasarnya agar makna inti dan representasi kalimat tetap terjaga. Setelah seluruh tahapan prapemrosesan selesai, data yang telah bersih ini kemudian digunakan sebagai basis untuk tahap labelling data dengan lexicon based.

### 2.3 Labeling Data

Setelah proses pembersihan data selesai, tahap berikutnya adalah pelabelan sentimen yang dilakukan secara otomatis menggunakan pendekatan lexicon-based[6]. Dalam metode ini, setiap komentar TikTok dianalisis berdasarkan kemunculan kata-kata yang telah dikategorikan dalam kamus sentimen. Komentar yang mengandung kata atau frasa bermuansa dukungan, apresiasi, atau pembelaan terhadap masyarakat adat dalam konflik agraria dengan PT Toba Pulp Lestari akan diberikan label 1 (positif). Sebaliknya, komentar yang memuat unsur tuduhan, kritik, atau sentimen negatif terhadap masyarakat adat akan diberi label 0 (negatif). Hasil dari proses ini adalah dataset siap pakai yang telah melalui tahap pembersihan dan pelabelan otomatis, sehingga dapat digunakan untuk tahap analisis selanjutnya, seperti tokenisasi, vektorisasi, dan pemodelan menggunakan Support Vector Machine.

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

TF-IDF (Term Frequency-Inverse Document Frequency) merupakan salah satu teknik penting dalam pemrosesan teks yang digunakan untuk menentukan seberapa signifikan suatu kata dalam sebuah koleksi komentar atau dokumen [7][8]. Pada penelitian ini, TF-IDF dimanfaatkan untuk menilai bobot setiap kata dalam kumpulan komentar TikTok terkait konflik agraria antara PT Toba Pulp Lestari dan masyarakat adat Toba Simalungun. Teknik ini efektif dalam mengurangi pengaruh kata-kata yang sering muncul di banyak komentar seperti kata-kata umum atau stopwords karena kata-kata tersebut umumnya tidak memberikan informasi yang relevan untuk analisis sentimen. Proses transformasi data teks menjadi bentuk numerik dilakukan menggunakan TfidfVectorizer, yang memungkinkan komentar-komentar tersebut diubah ke dalam format matriks angka sehingga dapat diolah lebih lanjut oleh algoritma klasifikasi, seperti Support Vector Machine [9][10]. Secara sederhana, term frequency (TF) menunjukkan seberapa sering suatu kata muncul dalam satu komentar, sementara inverse document frequency (IDF) mengukur seberapa jarang kata tersebut muncul di seluruh kumpulan data komentar. Dengan demikian, TF-IDF memberikan bobot lebih besar pada kata-kata yang spesifik dan bermakna dalam konteks komentar, sehingga menghasilkan representasi numerik yang lebih informatif untuk proses klasifikasi sentimen.

$$TF = \frac{f_{t,d}}{N_d} \quad (1)$$

$$Idf = \log \left( \frac{D+1}{df+1} \right) + 1 \quad (2)$$

$$W = TF * IDF \quad (3)$$

### 2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi yang banyak digunakan dalam berbagai penelitian, termasuk dalam analisis data teks. Berbasis pada prinsip statistik, SVM telah terbukti mampu memberikan hasil klasifikasi yang unggul dibandingkan dengan banyak teknik lain. Keunggulan utama dari SVM adalah kemampuannya dalam menangani data berdimensi tinggi, di mana teknik kernel yang digunakan memungkinkan pemetaan data dari ruang aslinya ke ruang berdimensi lebih tinggi sehingga pemisahan antar kelas menjadi lebih optimal[11]. Pada dasarnya, SVM bekerja dengan mencari hyperplane atau garis pembatas yang dapat memisahkan data ke dalam dua kategori, baik pada data yang bersifat linear maupun nonlinear. Kecepatan dan efisiensi SVM dalam proses klasifikasi juga menjadi salah satu alasan mengapa metode ini sering diadopsi dalam pengolahan data teks, seperti dalam penelitian analisis sentimen pada komentar TikTok.

### 2.6 Confusion Matrix

Confusion matrix merupakan sebuah tabel yang digunakan untuk menggambarkan performa hasil klasifikasi, dengan menunjukkan jumlah prediksi yang benar dan salah pada data uji[12]. Matriks ini berfungsi sebagai alat evaluasi dalam tugas klasifikasi machine learning, baik yang melibatkan dua kelas (biner) maupun lebih dari dua kelas (multikelas), sehingga memudahkan dalam mengukur tingkat akurasi model yang dikembangkan[13]. Pada penelitian ini, pengukuran kinerja model SVM dilakukan menggunakan sejumlah metrik evaluasi yang diperoleh dari confusion matrix, seperti precision, recall, accuracy, dan F1-score. Keempat metrik tersebut digunakan untuk menilai tingkat ketepatan model dalam mengklasifikasikan sentimen secara menyeluruh.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 1. Model Evaluasi SVM

## a. Accuracy

Accuracy merupakan perbandingan antara total prediksi yang benar baik prediksi positif maupun negative dengan jumlah seluruh data yang diuji. Rumusnya dituliskan sebagai:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

## b. Recall

Recall digunakan untuk menilai seberapa besar persentase data dengan label positif yang berhasil diidentifikasi dengan benar sebagai positif oleh model. Rumusnya adalah:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

## c. Precision

Precision menggambarkan tingkat ketepatan model dalam melakukan prediksi positif, yaitu persentase prediksi positif yang benar-benar sesuai atau relevan dengan label aslinya. Rumusnya adalah:

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+FN} \quad (6)$$

## d. F1-Score

F1-score adalah rata-rata harmonis antara nilai precision dan recall. Metrik ini sangat bermanfaat untuk menjaga keseimbangan antara keduanya, khususnya ketika jumlah data pada tiap kelas tidak seimbang.. Rumusnya:

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (7)$$

## Keterangan:

- *TP*: True Positives (data positif yang diprediksi positif)
- *TN*: True Negatives (data negatif yang diprediksi negatif)
- *FP*: False Positives (data negatif yang diprediksi positif)
- *FN*: False Negatives (data positif yang diprediksi negatif)

### 3. HASIL AND PEMBAHASAN

#### 3.1 Pengumpulan Data

Data dalam penelitian ini diperoleh dari platform media sosial TikTok, yang dikenal sangat dinamis dan mampu merepresentasikan opini publik secara real-time. TikTok dipilih karena menjadi salah satu media utama yang digunakan masyarakat Indonesia untuk mengekspresikan pandangan mengenai berbagai isu sosial, termasuk konflik agraria antara PT Toba Pulp Lestari dan masyarakat adat Toba Simalungun.

Pengambilan data dilakukan menggunakan API TikTok yang diintegrasikan dengan Python dan dijalankan pada lingkungan Google Colab. Untuk memastikan data yang diperoleh benar-benar relevan dengan fokus penelitian, peneliti menetapkan kata kunci utama yang berkaitan langsung dengan topik konflik agraria tersebut. Kata kunci ini dirancang untuk menangkap percakapan yang berkembang di masyarakat, baik secara eksplisit maupun implisit.

Periode pengambilan data dipilih pada saat isu konflik sedang ramai dibicarakan di TikTok, sehingga komentar yang dikumpulkan dapat mencerminkan dinamika opini publik yang aktual. Sebanyak 1.751 komentar berhasil dikumpulkan, masing-masing berupa teks mentah yang berisi opini atau respons pengguna terkait konflik tersebut.

Sebelum dianalisis lebih lanjut, seluruh data disimpan dalam format .csv dan hanya kolom teks komentar yang digunakan dalam proses analisis, sementara informasi tambahan seperti nama akun, waktu, atau ID komentar tidak diikutsertakan karena tidak relevan terhadap analisis sentimen berbasis konten. Pada tahap awal, data ini belum dilengkapi dengan label sentimen, sehingga selanjutnya dilakukan tahapan labelling data dengan text processing sehingga siap untuk dianalisis lebih lanjut.

Tabel 1. Sampel Dataset

No	Comment
1	DASAR ORANG ORANG TOBA ATAU BATAK INI TAK TAU MALU ITU TANAH SIMALUNGUN BUKAN TANAH BATAK.! ASAL MAIN KLAIM AJA KALIAN TOBA TOBA ITU GA TAU MALU KALIAN SEJAK KAPAN ADA TANAH ADAT TOBA ATAU BATAK DI TANAH SIMALUNGUNINI. !! WAJAR KALIAN DI PERLAKUKAN SEPERTI ITU KARNA KALIAN TAK TAU MALU . MENGATAS NAMAKAN TANAH LELUHUR KALIAN TANAH BATAK. SEMENTARA ITU TANAH SIMALUNGUN BUMI HABONARON DO BONA..! BUKAN TANAH BATAK..!! SEKALI LAGI TAU MALU DAN TAU DIRI LAH KALIAN PARA TOBA TOBA ATAU BATAK ITU JANGAN ASAL KLAIM TANAH WILAYAH KAMI SIMALUNGUN., mau lapor polisi, eh polisiny konco pt
2	kalo itu tanah negara berarti golongan hutan lindung dan apa kaitan pt tpl dengan lahan itu kalo benar ini pt tpl coba minta HGU mereka
3	biasanya gtu pemerintah daerah ada hubungannya.. bisa jadi dijual hutana
4	betul bg,semoga aparat pemerintah dapat bertindak tegas,karna kasus ini sudah lama saya lihat viral di tt ini,belum ada penyelesaiannya

### 3.2 Text Processing

Pengolahan data pada penelitian ini diawali dengan membersihkan teks mentah hasil ekstraksi komentar dari TikTok menggunakan platform Google Colab. Berbagai pustaka Python digunakan dalam tahap ini, antara lain re untuk merapikan karakter atau simbol yang tidak diperlukan, nltk untuk menghilangkan kata-kata umum (stopwords) yang kurang bermakna, serta Sastrawi sebagai alat stemming untuk mengembalikan kata-kata dalam Bahasa Indonesia ke bentuk dasarnya. Prosedur pembersihan ini sejalan dengan pendekatan yang diterapkan oleh [14], [15], yang berhasil mengelola ribuan data media sosial dalam penelitian klasifikasi sentimen.

Tahap awal pembersihan (cleaning) meliputi penghapusan tautan, emotikon, mention akun, tagar, angka, serta simbol-simbol lain yang tidak relevan untuk analisis opini. Seluruh teks kemudian dikonversi ke format huruf kecil (lowercase) guna menjaga konsistensi data. Daftar stopwords Bahasa Indonesia yang digunakan juga diperluas dengan berbagai kata informal yang sering dijumpai di media sosial, seperti "loh", "dong", atau "nih", agar proses pembersihan lebih optimal. Selanjutnya, dilakukan stemming untuk memastikan setiap kata kembali ke bentuk dasarnya sehingga makna utama dari komentar tetap terjaga dan akurat dalam analisis.

Setelah seluruh proses pembersihan selesai, data yang dihasilkan berupa kumpulan komentar yang telah siap untuk dianalisis lebih lanjut.

Tabel 2. Hasil Processing

Text Awal	Text Preprocessing
<p>DASAR ORANG ORANG TOBA ATAU BATAKINI TAK TAU MALU ITU TANAH SIMALUNGUN BUKAN TANAH BATAK! ASAL MAIN KLAIM AJA KALIAN TOBA TOBA ITU GA TAU MALU KALIAN SEJAK KAPAN ADA TANAH ADAT TOBA ATAU BATAK DI TANAH SIMALUNGUNINI. !! WAJAR KALIAN DI PERLAKUKAN SEPERTI ITU KARNA KALIAN TAK TAU MALU . MENGATAS NAMAKAN TANAH LELUHUR KALIAN TANAH BATAK. SEMENTARA ITU TANAH SIMALUNGUN BUMI HABONARON DO BONA..! BUKAN TANAH BATAK..!! SEKALI LAGI TAU MALU DAN TAU DIRI LAH KALIAN PARA TOBA TOBA ATAU BATAK ITU JANGAN ASAL KLAIM TANAH WILAYAH KAMI SIMALUNGUN.,</p> <p>mau lapor polisi, eh polisiny konco pt kalo itu tanah negara berarti golongan hutan lindung dan apa kaitan pt tpl dengan lahan itu kalo benar ini pt tpl coba minta HGU mereka biasanya gtu pemerintah daerah ada hubungannya.. bisa jadi dijual hutanaÃŸ betul bg,semoga aparat pemerintah dapat bertindak tegas,karna kasus ini sudah lama saya lihat viral di tt ini,belum ada penyelesaiannya</p>	<p>dasar orang orang toba batak tau malu tanah simalungun tanah batak main klaim toba toba tau malu tanah adat toba batak tanah simalungun wajar laku tau malu atas nama tanah leluhur tanah batak tanah simalungun bumi habonaron bona tanah batak tau malu tau toba batak klaim tanah wilayah simalungun</p> <p>lapor polisi polisiny konco pt tanah negara golong hutan lindung kait pt tpl lahan pt tpl coba hgu perintah daerah hubung jual hutana bgsemoga aparat perintah tindak tegaskarna lihat viral tt inibelum selesai</p>

### 3.3 Labelling Data

Setelah data komentar TikTok melalui tahap prapemrosesan, proses berikutnya adalah pelabelan sentimen pada masing-masing data. Proses labelling ini dilakukan secara otomatis menggunakan pendekatan lexicon-based yang telah diadaptasi untuk konteks konflik agraria antara PT Toba Pulp Lestari dan masyarakat adat Toba Simalungun. Setiap komentar dianalisis berdasarkan kemunculan kata atau frasa tertentu yang telah diklasifikasikan ke dalam kategori sentimen positif atau negatif.

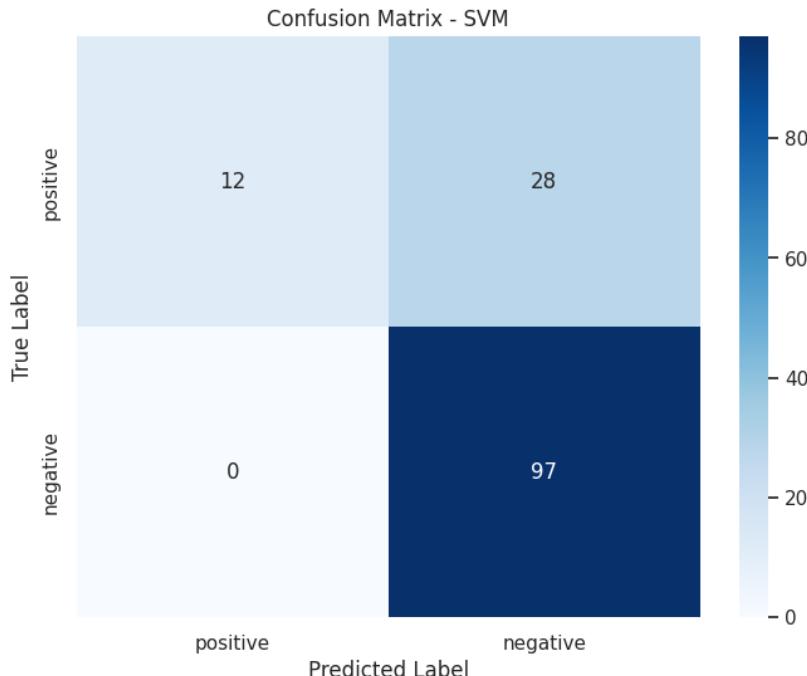
Komentar yang dinilai memberikan dukungan, empati, atau pembelaan terhadap masyarakat adat dikelompokkan sebagai sentimen positif, sedangkan komentar yang mengandung kritik, tuduhan, atau kecenderungan mendukung perusahaan dikelompokkan sebagai sentimen negatif. Proses ini menghasilkan dua label utama, yaitu 1 untuk sentimen positif dan 0 untuk sentimen negatif. Hasil pelabelan ini dapat dilihat pada Tabel 4.1 berikut, yang menampilkan distribusi jumlah komentar pada masing-masing kategori sentimen.

Tabel 3. Hasil labelling Data

Text Preprocessing	Sentiment
dasar orang orang toba batak tau malu tanah simalungun tanah batak main klaim toba toba tau malu tanah adat toba batak tanah simalungun wajar laku tau malu atas nama tanah leluhur tanah batak tanah simalungun bumi habonaron bona tanah batak tau malu tau toba batak klaim tanah wilayah simalungun	0
lapor polisi polisiny konco pt	0
tanah negara golong hutan lindung kait pt tpl lahan pt tpl coba hgu	1
perintah daerah hubung jual hutana	0
bgsemoga aparat perintah tindak tegaskarna lihat viral tt inibelum selesai	0

### 3.4 Evaluasi Model

Evaluasi performa model Support Vector Machine (SVM) dalam mengklasifikasikan sentimen publik terhadap konflik agraria antara PT Toba Pulp Lestari dan masyarakat adat Toba Simalungun dilakukan dengan menggunakan confusion matrix serta metrik-metrik evaluasi seperti accuracy, precision, recall, dan F1-score. Gambar 1 berikut memperlihatkan confusion matrix hasil prediksi model terhadap data uji:



Gambar 2. Confusion Matrix

Berdasarkan Gambar 1, dapat dijelaskan bahwa dari total 137 data uji, model SVM berhasil mengklasifikasikan 97 data dengan label negatif secara benar (true negative). Untuk data dengan label positif, terdapat 12 data yang dapat diprediksi dengan tepat sebagai positif (true positive), namun masih terdapat 28 data positif yang salah terkласifikasi sebagai negatif (false negative). Menariknya, pada hasil ini tidak ditemukan adanya data negatif yang salah terkласifikasi sebagai positif (false positive). Hal ini menunjukkan bahwa model sangat baik dalam mengenali sentimen negatif, namun masih cukup banyak komentar positif yang tidak terdeteksi dengan baik oleh model.

Hasil pengukuran accuracy, precision, recall, dan F1-score untuk setiap kelas sentimen ditampilkan pada Tabel 3 berikut.

Tabel 4. Hasil Evaluasi Model SVM

Label	Precision	Recall	F1-Score	Support
Negatif	0.78	1.00	0.87	97
Positif	1.00	0.30	0.46	40
Accuracy			0.80	137
Macro Avg	0.89	0.65	0.67	137
Weighted Avg	0.84	0.80	0.75	137

Berdasarkan tabel di atas, nilai **accuracy** model adalah 79,56%, yang menunjukkan bahwa sekitar 80% data uji berhasil diklasifikasikan dengan tepat. Precision untuk kelas negatif mencapai 0.78, dengan recall yang sangat tinggi yaitu 1.00, menandakan hampir seluruh komentar negatif dapat dikenali oleh model dengan baik. Namun demikian, pada kelas positif, meskipun precision sangat tinggi (1.00), nilai recall-nya rendah (0.30). Hal ini berarti sebagian besar komentar positif justru tidak teridentifikasi sebagai positif, dan cenderung diklasifikasikan sebagai negatif. F1-score pada kelas negatif juga cukup tinggi (0.87), namun untuk kelas positif hanya sebesar 0.46.

Ketimpangan antara performa pada kelas positif dan negatif ini dapat disebabkan oleh distribusi data yang tidak seimbang, atau karena ciri-ciri linguistik sentimen positif dalam komentar terkait konflik agraria relatif lebih beragam atau tidak eksplisit dibanding sentimen negatif. Fenomena ini juga ditemukan pada penelitian analisis sentimen berbasis machine learning di ranah sosial politik [16][17], di mana sentimen negatif cenderung lebih mudah dikenali model karena sering kali memiliki kata-kata yang lebih tegas dan eksplisit.

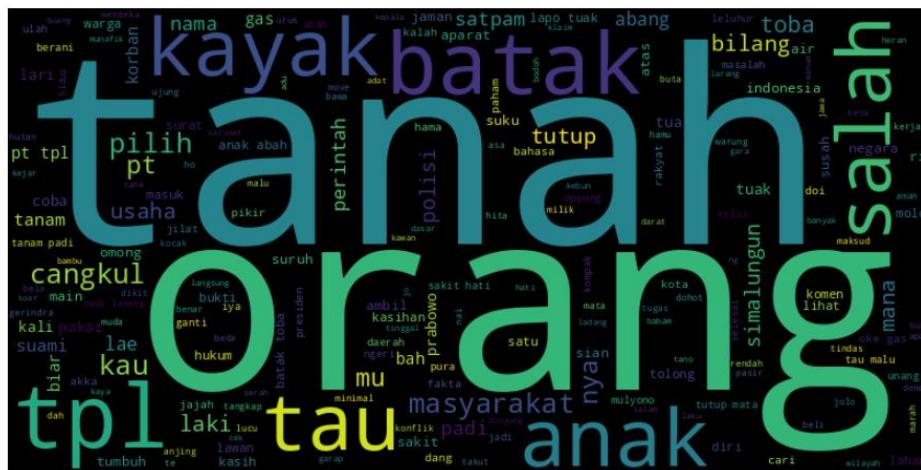
Secara umum, performa model SVM dalam penelitian ini sudah cukup baik dalam mengidentifikasi sentimen negatif pada komentar TikTok tentang konflik agraria PT Toba Pulp Lestari dan masyarakat adat Toba Simalungun. Namun, diperlukan perbaikan atau penyesuaian lebih lanjut misalnya pada tahap preprocessing, penyeimbangan data, atau penambahan fitur linguistik agar kemampuan model dalam mendeteksi sentimen positif dapat meningkat. Temuan ini sejalan dengan literatur sebelumnya yang menyatakan bahwa SVM sangat efektif untuk data berdimensi tinggi, namun sangat dipengaruhi oleh kualitas fitur dan distribusi label [18][19].

Dengan hasil evaluasi ini, penelitian dapat memberikan gambaran objektif terkait pola sentimen publik di TikTok, serta memperkuat peran machine learning dalam riset sosial berbasis media digital di Indonesia.

### 3.5 Visualisasi Wordcloud Sentimen

Setelah evaluasi kuantitatif terhadap performa model *Support Vector Machine* (SVM) melalui analisis *confusion matrix* dan metrik-metrik klasifikasi relevan (seperti akurasi, presisi, dan *recall*), penelitian dilanjutkan ke tahap analisis kualitatif. Tahap ini memanfaatkan teknik visualisasi *wordcloud* untuk mendapatkan pemahaman yang lebih mendalam mengenai komposisi leksikal pada setiap kategori sentimen. *Wordcloud* tersebut berfungsi sebagai representasi grafis yang menyoroti frekuensi dan dominansi istilah-istilah kunci dalam korpus komentar TikTok, yang secara spesifik membahas konflik agraria antara PT Toba Pulp Lestari dengan masyarakat adat Toba Simalungun, baik dari perspektif sentimen positif maupun negatif.

Pada **Gambar 2** ditampilkan wordcloud untuk kategori sentimen negatif. Kata-kata seperti “tanah”, “orang”, “salah”, “tutupan”, “toba”, “batak”, dan “simalungun” mendominasi visualisasi ini. Kemunculan kata “tanah” sebagai kata yang paling menonjol mengindikasikan bahwa perdebatan kepemilikan lahan menjadi isu utama yang menimbulkan sentimen negatif di tengah masyarakat. Istilah seperti “salah”, “tutupan”, serta “tau malu” merefleksikan adanya narasi kecaman, kritik, atau tuduhan yang diarahkan baik kepada individu, kelompok, maupun institusi terkait. Kata-kata seperti “polisi”, “perintah”, dan “pt tpl” juga muncul, menandakan banyak komentar negatif yang mengkritik peran pemerintah atau pihak perusahaan dalam konflik tersebut. Selain itu, ditemukan juga kosakata yang berasal dari masyarakat lokal seperti “toba”, “batak”, “simalungun”, dan “lae”, menandakan keterlibatan dan identitas etnis sangat berpengaruh dalam diskusi publik mengenai isu ini.



Gambar 3. Wordcloud Negatif SVM

Sementara itu, Gambar 3 memperlihatkan wordcloud untuk kategori sentimen positif. Pada wordcloud ini, kata “tanah” tetap menjadi kata yang paling sering muncul, diikuti oleh kata-kata seperti “bujang”, “moga”, “pulp”, “lestari”, “miliki”, “lihat”, dan “polisi”. Keberadaan kata “bujang” dan “moga” merefleksikan harapan, doa, dan dukungan yang disampaikan oleh pengguna kepada masyarakat adat atau kelompok yang terlibat. Istilah “pulp”, “lestari”, dan “pt tpl” menunjukkan adanya diskusi terkait perusahaan, namun dalam konteks yang cenderung lebih netral atau mendukung resolusi damai. Kata “miliki”, “kerja”, dan “usaha” menandakan adanya narasi tentang hak kepemilikan tanah dan dorongan untuk menemukan solusi yang adil serta keberlanjutan ekonomi bagi masyarakat terdampak.



Gambar 4. Wordcloud Positif SVM

Analisis wordcloud ini memperkuat hasil evaluasi model yang telah dijabarkan sebelumnya. Dominasi kata-kata dengan muatan negatif pada kategori negatif selaras dengan hasil confusion matrix, di mana model SVM sangat akurat dalam mendeteksi komentar bernuansa negatif. Hal ini didukung pula oleh penelitian [20] dan [21], yang menunjukkan bahwa kata-kata bernuansa kritik, tuduhan, atau kecaman lebih mudah dikenali oleh algoritma klasifikasi berbasis machine learning karena memiliki pola linguistik yang lebih eksplisit dan tegas. Sebaliknya, pada kategori positif, ragam kata lebih bervariasi dan terkadang kurang eksplisit dalam mengekspresikan dukungan, sehingga model mengalami penurunan recall pada kelas ini.

Selain itu, temuan dari wordcloud juga menunjukkan adanya keterkaitan erat antara identitas lokal, struktur kekuasaan, dan dinamika konflik, yang menjadi faktor penting dalam analisis sentimen sosial berbasis media digital. Visualisasi ini dapat menjadi dasar dalam pengambilan keputusan bagi stakeholder terkait, seperti pemerintah, perusahaan, maupun organisasi masyarakat sipil, untuk memahami persepsi publik dan merancang strategi komunikasi yang lebih efektif dalam upaya penyelesaian konflik agraria.

Dengan demikian, integrasi antara evaluasi numerik dan analisis visual wordcloud memberikan gambaran yang komprehensif mengenai pola sentimen publik di media sosial TikTok terhadap isu agraria yang diangkat dalam penelitian ini.

#### 4. KESIMPULAN

Penelitian ini menunjukkan bahwa analisis sentimen publik terhadap konflik agraria antara PT Toba Pulp Lestari dan masyarakat adat Toba Simalungun dapat dilakukan secara efektif dengan memanfaatkan metode Support Vector Machine (SVM) berbasis data komentar TikTok. Rangkaian proses yang meliputi pengumpulan data, prapemrosesan teks, pelabelan sentimen otomatis, transformasi fitur menggunakan TF-IDF, hingga klasifikasi dengan SVM, menghasilkan model yang mampu membedakan sentimen negatif dan positif secara cukup akurat. Evaluasi model memperlihatkan bahwa SVM memiliki akurasi sebesar 80%, dengan performa yang sangat baik dalam mendeteksi komentar bernada negatif, meskipun pengenalan terhadap sentimen positif masih memerlukan penguanan lebih lanjut.

Keterhubungan antara tujuan penelitian yang menekankan pemetaan opini publik secara objektif dan hasil analisis yang diuraikan dalam bab hasil dan pembahasan menegaskan relevansi pendekatan machine learning dalam riset sosial berbasis data media sosial. Visualisasi wordcloud turut memperkaya interpretasi hasil dengan mengidentifikasi kosakata utama yang mencerminkan karakteristik tiap sentimen, sekaligus menyoroti dimensi identitas lokal dan dinamika konflik yang terekspresi di ruang digital.

Prospek pengembangan dari penelitian ini terbuka luas, di antaranya melalui perluasan analisis ke berbagai platform media sosial lain, penerapan teknik penyeimbangan data untuk mengatasi distribusi kelas yang tidak seimbang, serta eksplorasi fitur kontekstual dan semantik guna meningkatkan kinerja klasifikasi, khususnya pada sentimen positif. Penerapan metode serupa juga dapat diarahkan untuk menganalisis isu-isu sosial lainnya, sehingga mampu berkontribusi pada penyusunan kebijakan berbasis data dan penguatan advokasi publik. Pengembangan lebih lanjut juga dapat difokuskan pada pembuatan sistem pemantauan sentimen publik secara real-time atau integrasi dengan data spasial untuk mendukung penyelesaian konflik agraria di tingkat yang lebih luas.

Dengan demikian, temuan penelitian ini berkontribusi dalam memperluas literatur mengenai analisis sentimen sosial berbasis machine learning di Indonesia, serta membuka peluang bagi pengembangan inovasi di bidang pemantauan opini publik dan pengelolaan isu-isu strategis di era digital.

#### REFERENCES

- [1] A. Alfian, K. Kudussisara, N. S. Maimunah, and I. Susana, "Masyarakat Adat dan Toba PULP Lestari: Pemetaan Aktor dan Analisis Konflik Agraria," *Aceh Anthropological Journal*, vol. 9, no. 1, pp. 53–69, May 2025, doi: 10.29103/aaaj.v9i1.20332.
- [2] F. Tobing, "SENGKETA TANAH ANTARA MASYARAKAT ADAT BATAK DENGAN PT. TOBA PULP LESTARI DAN PELANGGARAN PERBUATAN-PERBUATAN YANG MENCIDERAI ATURAN KEHUTANAN DI WILAYAH SUMATERA UTARA," *Jurnal Ilmu Hukum Sui Generis*, vol. 2, no. 2, pp. 77–81, 2022, doi: 10.23887/jih.v2i2.1014.
- [3] N. Norlaila, W. W. Winarno, and E. T. Luthfi, "Analisis Sentimen Masyarakat Tentang Tambang Di Indonesia Pada Twitter Menggunakan Data Mining," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, pp. 1091–1099, Aug. 2024, doi: 10.29100/jipi.v9i3.5402.
- [4] R. Darman, "ANALISIS SENTIMEN RESPON TWITTER TERHADAP PERSYARATAN BADAN PENYELENGGARA JAMINAN SOSIAL (BPJS) DI KANTOR PERTANAHAN," *JURNAL WIDYA BHUMI*, vol. 3, pp. 113–136, 2023, doi: 10.31292/wb.v3i2.61.
- [5] D. Ratna Rohmania and R. Abidin, "LogicLink : Journal of Artificial Intelligence and Multimedia in Informatics Analisis Sentimen Masyarakat Indonesia Terhadap Kebijakan Program Tapera Menggunakan Brand24," *LogicLink : Journal of Artificial Intelligence and Multimedia in Informatics*, vol. 1, no. 2, pp. 120–131, 2024, doi: 10.28918/logiclink.v1i2.8736.
- [6] A. P. Putra and A. F. Syafira, "Analisis Sentimen Data Twitter Topik Politik Dengan Metode Naive Bayes Dan Convolutional Neural Networks (Cnn)," *Jurnal Ilmiah Wahana Pendidikan, Oktober*, vol. 9, no. 20, pp. 36–41, 2023, doi: 10.5281/zenodo.8396579.
- [7] K. X. Han, W. Chien, C. C. Chiu, and Y. T. Che ng, "Application of support vector machine (SVM) in the sentiment analysis of twitter dataset," *Applied Sciences (Switzerland)*, vol. 10, no. 3, Feb. 2020, doi: 10.3390/app10031125.
- [8] F. Es-sabery, I. Es-sabery, J. Qadir, B. Sainz-de-Abajo, and B. Garcia-Zapirain, "A hybrid Hadoop-based sentiment analysis classifier for tweets associated with COVID-19 utilizing two machine learning algorithms: CNN, and fuzzy C4.5," *J Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-01014-4.

- [9] I. M. Parapat, M. T. Furqon, and Sutrisno, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, pp. 3163–3169, 2018.
- [10] M. Muchtar and R. A. Muchtar, "PERBANDINGAN METODE KNN DAN SVM DALAM KLASIFIKASI KEMATANGAN BUAH MANGGA BERDASARKAN CITRA HSV DAN FITUR STATISTIK," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4010.
- [11] F. C. Onwuegbuche, J. M. Wafula, and J. K. Mung'atu, "Support Vector Machine for Sentiment Analysis of Nigerian Banks Financial Tweets," *Journal of Data Analysis and Information Processing*, vol. 07, no. 04, pp. 153–173, 2019, doi: 10.4236/jdaip.2019.74010.
- [12] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 6, no. 3, pp. 150–55, 2021, doi: 10.30591/jpit.v6i3.2870.
- [13] M. R. Fahlevvi, "Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di Google Playstore Menggunakan Metode Support Vector Machine," *Jurnal Teknologi dan Komunikasi Pemerintahan*, vol. 4, no. 1, pp. 1–13, 2022, [Online]. Available: <http://ejournal.ipdn.ac.id/JTKP>
- [14] V. Fitriyana *et al.*, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," 2023.
- [15] S. D. Simamora, F. Irwiensyah, and F. N. Hasan, "Analisis Sentimen Terkait Konflik Palestina-Israel Pada Media Sosial X Menggunakan Algoritma Naïve Bayes Classifier," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 1, Jun. 2024, doi: 10.47065/bits.v6i1.5344.
- [16] M. Dehghani and Z. Yazdanparast, "Political Sentiment Analysis of Persian Tweets Using CNN-LSTM Model," 2023. doi: 10.48550/arXiv.2307.07740.
- [17] G. Z. Rahma and M. Rosyda, "Perbandingan Kinerja Algoritma Convolutional Neural Network (CNN) dan Recurrent Neural Network (RNN) pada Analisis Sentimen Pemilu Presiden 2024," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 2, Feb. 2025, doi: 10.52436/1.jpti.652.
- [18] M. Hadi Arfian *et al.*, "Analisis Sentimen Pada Media Sosial Menggunakan Metode Support Vector Machine," *Jurnal Ilmu Teknik dan Komputer*, vol. 09, no. 01, pp. 1–6, 2025, doi: 10.22441/jitkom.v9i1.001.
- [19] A. Liawati, R. Narasati, D. Solihudin, C. Lukman Rohmat, and S. Eka Permana, "ANALISIS SENTIMEN KOMENTAR POLITIK DI MEDIA SOSIAL X DENGAN PENDEKATAAN DEEP LEARNING," 2023.
- [20] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, "PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER," *JOISM : JURNAL OF INFORMATION SYSTEM MANAGEMENT*, vol. 3, no. 1, pp. 16–1, 2021, doi: 10.24076/joism.2021v3i2.558.
- [21] M. Rahardi, A. Aminuddin, F. F. Abdullah, and R. A. Nugroho, "Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, pp. 534–539, 2022, doi: 10.14569/IJACSA.2022.0130665.