

ANALISIS SENTIMEN TERHADAP PROGRAM KOPERASI DESA MERAH PUTIH DI MULTI PLATFORM MENGGUNAKAN METODE TF-IDF DAN SVM

Lailatul Qodriyah, Ferdita Inayah Hestu Saputri, M. Ahsanul Fikri, Ifnu Wisma Dwi Prastyia

Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri
Jalan Jendral Ahmad Yani No. 10, Bojonegoro, Jawa Timur, Indonesia
ahsanulfikri26@gmail.com

ABSTRAK

Program Koperasi Desa Merah Putih merupakan kebijakan strategis pemerintah dalam meningkatkan perekonomian desa melalui koperasi berbasis gotong royong. Seiring berkembangnya media digital, masyarakat secara aktif menyampaikan opini terhadap program ini melalui berbagai platform, sehingga menghasilkan data teks tidak terstruktur yang sulit dianalisis secara manual dan belum menggambarkan persepsi publik secara objektif. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Program Koperasi Desa Merah Putih secara multi-platform guna memperoleh gambaran opini publik yang komprehensif. Metode yang digunakan adalah text mining dengan ekstraksi fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) dan klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM). Data komentar dikumpulkan dari platform X (Twitter) dan TikTok melalui teknik web scraping, kemudian dilakukan preprocessing, pelabelan sentimen, serta evaluasi model menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan adanya perbedaan karakteristik sentimen antar platform, di mana sentimen positif mendominasi pada platform X, sedangkan sentimen negatif lebih banyak ditemukan pada platform TikTok. Model SVM menghasilkan akurasi sebesar 98,73% pada platform X dan 94,32% pada platform TikTok, yang menunjukkan bahwa kombinasi TF-IDF dan SVM efektif dalam mengklasifikasikan sentimen masyarakat terhadap kebijakan publik.

Kata kunci : Analisis Sentimen, Koperasi Desa Merah Putih, TF-IDF, Support Vector Machine, Media Sosial

1. PENDAHULUAN

Koperasi Desa Merah Putih (KDMP) adalah program strategis pemerintah yang bertujuan untuk meningkatkan ekonomi lokal di tingkat desa melalui pendekatan koperasi kekeluargaan dan gotong royong [1]. Konsep ini muncul sebagai tanggapan terhadap kebutuhan masyarakat desa akan lembaga ekonomi yang dapat meningkatkan ketahanan pangan, mendukung pertumbuhan usaha mikro, dan memberikan kontribusi berkelanjutan terhadap kesejahteraan warga. Melalui keterlibatan masyarakat dalam sistem usaha bersama, KDMP membantu memperkuat struktur ekonomi desa dengan mengelola unit usaha yang sesuai dengan potensi lokal.

Namun, besarnya volume opini masyarakat yang tersebar di berbagai platform digital seperti Twitter (X), TikTok, dan media online menghasilkan data teks yang tidak terstruktur dan sulit dianalisis secara manual. Kondisi ini menyebabkan persepsi masyarakat terhadap Program Koperasi Desa Merah Putih belum dapat dipetakan secara objektif dan menyeluruh.

Selain itu, hasil analisis sentimen sangat dipengaruhi oleh proses representasi fitur teks dan pemilihan algoritma klasifikasi yang digunakan. Representasi fitur yang kurang tepat dapat menurunkan akurasi model, sehingga diperlukan metode ekstraksi fitur yang efektif serta algoritma klasifikasi yang mampu menangani data teks berdimensi tinggi.

Di sisi lain, meskipun analisis sentimen terhadap kebijakan publik telah banyak dilakukan, penelitian

yang secara khusus menganalisis sentimen masyarakat terhadap Program Koperasi Desa Merah Putih dengan memanfaatkan data dari berbagai platform digital (multi-platform) masih sangat terbatas. Oleh karena itu, diperlukan penelitian yang mampu mengintegrasikan teknik text mining dan pembelajaran mesin untuk mengklasifikasikan sentimen masyarakat secara akurat dan komprehensif..

Teknik text mining menjadi pendekatan penting dalam pengolahan data berbasis teks untuk menggali informasi dari data tidak terstruktur [2]. Analisis sentimen adalah metode yang banyak digunakan untuk menempatkan pendapat atau perspektif yang ada dalam teks ke dalam kategori seperti positif, negatif, atau netral [3]. Metode ini memungkinkan pengolahan opini masyarakat secara otomatis dalam skala besar. Ini dapat memberikan gambaran umum tentang bagaimana masyarakat melihat program atau kebijakan pemerintah. Tetapi proses representasi fitur teks dan pemilihan algoritma klasifikasi yang tepat sangat memengaruhi akurasi hasil analisis sentimen.

Metode TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen serta tingkat kepentingannya terhadap keseluruhan dokumen [4]. TF-IDF cocok untuk data teks berskala besar dari berbagai platform digital karena mampu merepresentasikan teks ke dalam bentuk numerik yang efektif dan relatif sederhana. Dalam analisis sentimen, kinerja algoritma klasifikasi dipengaruhi langsung oleh representasi fitur yang baik. Untuk proses klasifikasi sentimen, Support Vector Machine (SVM) merupakan salah satu algoritma

supervised learning yang sering digunakan karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan performa klasifikasi yang baik [5]. Hyperplane optimal dicari oleh SVM untuk memisahkan data ke dalam kelas tertentu dengan margin maksimum. Studi sebelumnya menunjukkan bahwa kombinasi TF-IDF dan SVM sangat akurat dalam analisis sentimen pada data media sosial dan ulasan online [6]. Karena keunggulan ini, SVM menjadi algoritma yang layak untuk digunakan dalam penelitian berbasis opini publik.

Studi sebelumnya telah mengkaji analisis sentimen terhadap kebijakan publik dan program pemerintah dengan menggunakan berbagai teknik pembelajaran mesin. Beberapa penelitian mengklasifikasikan sentimen masyarakat dengan menggunakan Naive Bayes, k-Nearest Neighbor, dan Decision Tree [7]. Hasilnya, bagaimanapun, menunjukkan bahwa karakteristik data dan teknik ekstraksi fitur yang digunakan sangat memengaruhi kinerja model. Menurut penelitian tambahan, SVM biasanya lebih stabil dan unggul dalam menangani data teks yang kompleks dan tidak seimbang. Namun, penelitian analisis sentimen yang secara khusus memeriksa Program Koperasi Desa Merah Putih menggunakan pendekatan multi platform masih sangat sedikit.

Berdasarkan masalah tersebut, penelitian ini bertujuan untuk menganalisis persepsi masyarakat terhadap Program Koperasi Desa Merah Putih pada berbagai platform digital. Penelitian ini menggunakan metode ekstraksi fitur TF-IDF dan algoritma klasifikasi Support Vector Machine. Untuk mendapatkan gambaran sentimen yang lebih lengkap, data dikumpulkan dari berbagai platform digital. Penelitian ini diharapkan dapat memberikan kontribusi akademis untuk pengembangan penelitian analisis sentimen berbasis text mining dan machine learning serta memberikan wawasan bagi pemerintah dan pemangku kebijakan dalam mengevaluasi implementasi program.

2. TINJAUAN PUSTAKA

Analisis sentimen banyak digunakan untuk memahami pendapat masyarakat tentang produk, layanan, atau kebijakan publik berdasarkan data teks yang dikumpulkan dari berbagai platform digital. Proses identifikasi polaritas dalam data dokumen atau teks menghasilkan klasifikasi emosional menjadi sentimen positif, negatif, atau netral.

Salah satu sumber utama data opini publik adalah media digital, yang memungkinkan pengguna menyampaikan pengalaman dan pandangan mereka secara terbuka. Dengan menggunakan platform digital seperti media sosial dan media online, orang dapat memberikan umpan balik, kritik, dan dukungan terhadap program atau kebijakan. Untuk menganalisis data opini yang efektif, teknik pengolahan data berbasis komputasi diperlukan karena data opini yang

dihasilkan dari platform tersebut biasanya berbentuk teks tidak terstruktur dan berukuran besar.

Scraping web adalah metode otomatis dan terstruktur untuk mengumpulkan data dari situs web tertentu[8]. Ini memungkinkan data untuk disimpan dalam format yang dapat diproses tanpa perlu dilakukan secara manual oleh pengguna. Ini banyak digunakan dalam penelitian berbasis text mining untuk mengumpulkan data opini dari berbagai platform digital.

Text mining adalah proses pengolahan data teks tidak terstruktur yang bertujuan untuk mengekstraksi pola dan informasi tertentu[9]. Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode ekstraksi fitur yang umum digunakan karena mampu memberikan bobot pada kata berdasarkan tingkat kepentingannya dalam dokumen dan keseluruhan korpus[10].

Salah satu metode klasifikasi supervised learning adalah Support Vector Machine (SVM), yang bekerja dengan menemukan hyperplane yang ideal untuk memisahkan data ke dalam kelas tertentu dengan margin maksimum. SVM banyak digunakan dalam analisis sentimen karena memiliki keunggulan dalam menangani data teks yang kompleks dan besar [11] [12]. Dalam sejumlah penelitian analisis sentimen berbasis teks, kombinasi TF-IDF dan SVM menunjukkan kinerja klasifikasi yang stabil dan baik.

Studi sebelumnya menunjukkan bahwa TF-IDF sebagai representasi fitur dan SVM sebagai algoritma klasifikasi dapat mengungguli metode lain seperti Naive Bayes [13]. Ini terutama berlaku untuk data teks dengan distribusi kelas yang tidak seimbang dan variasi kata yang besar. Untuk menganalisis sentimen masyarakat terhadap Program Koperasi Desa Merah Putih secara objektif dan akurat, pendekatan TF-IDF dan SVM digunakan dalam penelitian ini.

Penelitian terdahulu telah menerapkan analisis sentimen kdm pada platform Twitter (X) menggunakan algoritma SVM dan Naive Bayes. Hasil penelitian tersebut menunjukkan bahwa SVM umumnya menghasilkan tingkat akurasi yang lebih tinggi dibandingkan Naive Bayes, terutama pada data teks dengan variasi kata yang tinggi dan distribusi kelas yang tidak seimbang [14]. Namun, sebagian besar penelitian tersebut hanya memanfaatkan satu platform media sosial, sehingga cakupan persepsi masyarakat yang dianalisis masih terbatas.

Berbeda dengan penelitian sebelumnya, penelitian ini menggunakan pendekatan multi-platform dengan menggabungkan data opini masyarakat dari Twitter (X) dan TikTok. Data teks yang diperoleh direpresentasikan menggunakan metode TF-IDF dan diklasifikasikan menggunakan algoritma Support Vector Machine (SVM). Pendekatan ini diharapkan mampu memberikan gambaran sentimen masyarakat yang lebih komprehensif dan akurat terhadap Program Koperasi Desa Merah Putih dibandingkan penelitian yang hanya menggunakan satu sumber data.

3. METODE PENELITIAN

3.1. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode text mining, analisis sentimen adalah salah satu bagian dari text mining [15]. Pendekatan kuantitatif dipilih karena penelitian ini berfokus pada pengolahan data teks dalam jumlah besar yang kemudian direpresentasikan dalam bentuk numerik dan dianalisis menggunakan algoritma pembelajaran mesin. Analisis sentimen dilakukan untuk mengetahui kecenderungan opini masyarakat terhadap Program Koperasi Desa Merah Putih berdasarkan komentar pengguna pada berbagai platform digital.

3.2. Objek Penelitian

Objek penelitian adalah komentar masyarakat yang membahas Program Koperasi Desa Merah Putih pada platform X (Twitter) dan TikTok. Data berupa teks komentar yang bersifat tidak terstruktur dan mengandung opini publik, sehingga memerlukan proses pengolahan data sebelum dilakukan klasifikasi sentimen.

3.3. Sumber Data dan Teknik Pengumpulan Data

Sumber data dalam penelitian ini berasal dari platform digital yang bersifat publik. Teknik pengumpulan data dilakukan menggunakan metode web scraping, dengan penyesuaian terhadap karakteristik masing-masing platform, yaitu sebagai berikut:

a. Platform X (Twitter)

Pengumpulan data komentar dilakukan menggunakan library Snsrcape pada lingkungan Google Colab [16]. Proses scraping dilakukan dengan menentukan kata kunci yang relevan dengan topik Program Koperasi Desa Merah Putih. Data yang dikumpulkan berupa teks komentar/tweet yang kemudian disimpan dalam format terstruktur untuk tahap pengolahan selanjutnya.

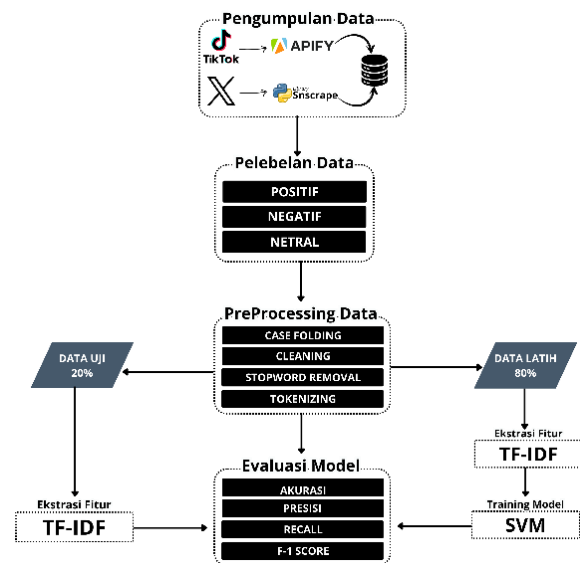
b. Platform TikTok

Data komentar TikTok diperoleh menggunakan tools Apify [17]. Proses scraping dilakukan dengan memanfaatkan tautan video atau konten yang berkaitan dengan Program Koperasi Desa Merah Putih. Komentar hasil scraping diekspor ke dalam format CSV atau Excel agar dapat diproses lebih lanjut pada tahap preprocessing dan analisis.

3.4. Alur Penelitian

Tahapan penelitian ini disusun secara sistematis untuk menggambarkan proses penelitian dari awal hingga akhir. Alur penelitian dimulai dari pengumpulan data komentar dari platform X (Twitter) dan TikTok, dilanjutkan dengan preprocessing data teks, pelabelan sentimen, ekstraksi fitur menggunakan TF-IDF, proses klasifikasi menggunakan algoritma Support Vector Machine (SVM), hingga evaluasi

performa model. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian

3.5. Preprocessing Data Teks

Data komentar yang diperoleh dari hasil scraping masih mengandung noise seperti tanda baca, URL, emoji, angka, serta variasi huruf besar dan kecil [18]. Oleh karena itu, dilakukan tahap preprocessing untuk meningkatkan kualitas data. Tahapan preprocessing yang diterapkan dalam penelitian ini meliputi:

- Case folding, yaitu mengubah seluruh teks menjadi huruf kecil.
- Cleaning, untuk menghapus URL, simbol, emoji, angka, dan karakter tidak relevan.
- Tokenizing, yaitu memecah teks menjadi kata-kata penyusun.
- Stopword removal, untuk menghilangkan kata-kata umum yang tidak memiliki makna sentimen signifikan.

Hasil dari preprocessing ini berupa teks bersih (*Clean_Comment*) yang siap digunakan pada tahap pelabelan dan pemodelan.

3.6. Pelabelan Data Sentimen

Pelabelan sentimen pada penelitian ini dilakukan secara otomatis menggunakan pendekatan rule-based berbasis kata kunci [19]. Setiap komentar dianalisis dengan mencocokkan kata-kata yang terdapat di dalamnya terhadap daftar kata kunci sentimen positif. Komentar yang mengandung kata kunci positif diberi label positif, sedangkan komentar lainnya diberi label negatif.

Pendekatan ini digunakan untuk mempercepat proses pelabelan data dalam jumlah besar serta menjaga konsistensi penentuan kelas sentimen. Hasil pelabelan disimpan dalam kolom *Label* dan digunakan sebagai target kelas pada proses klasifikasi.

3.7. Ekstraksi Fitur Menggunakan TF-IDF

Setelah pelabelan, data teks diubah ke dalam bentuk numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam satu dokumen dan tingkat kepentingannya terhadap keseluruhan korpus. Representasi TF-IDF dipilih karena efektif dalam menangani data teks berdimensi tinggi dan sering digunakan dalam analisis sentiment [20].

3.8. Pembagian Data Latih dan Data Uji

Dataset yang telah diekstraksi fiturnya dibagi menjadi dua bagian, yaitu 80% data latih dan 20% data uji. Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengukur kemampuan model dalam mengklasifikasikan data baru yang belum pernah dilihat sebelumnya [21].

3.9. Klasifikasi Menggunakan Support Vector Machine (SVM)

Proses klasifikasi sentimen dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dengan pendekatan linear, yang diimplementasikan menggunakan *LinearSVC* dari pustaka *scikit-learn*. Algoritma ini bekerja dengan mencari hyperplane optimal yang mampu memisahkan dua kelas sentimen dengan margin maksimum [22]. SVM linear dipilih karena sesuai untuk data teks berbasis TF-IDF yang memiliki dimensi tinggi dan bersifat sparse.

3.10. Evaluasi Model

Evaluasi performa model dilakukan menggunakan beberapa metrik, yaitu akurasi, precision, recall, F1-score, serta confusion matrix [23]. Evaluasi ini bertujuan untuk mengetahui tingkat keberhasilan model dalam mengklasifikasikan sentimen komentar masyarakat terhadap Program Koperasi Desa Merah Putih.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pada tahap awal penelitian ini, data komentar masyarakat tentang Program Koperasi Desa Merah Putih dikumpulkan dari dua platform digital, X (Twitter) dan TikTok. Data dikumpulkan menggunakan teknik web scraping dengan mempertimbangkan fitur masing-masing platform. Pada platform X, data dikumpulkan menggunakan library *Snsrape* melalui Google Colab dengan kata kunci yang relevan dengan topik penelitian. Sementara itu, alat *Apify* digunakan untuk mengumpulkan data komentar dari TikTok dengan menghubungkan tautan ke konten video yang membahas Program Koperasi Desa Merah Putih.

Data yang dikumpulkan dari kedua platform tidak terstruktur dan berisi berbagai jenis opini

masyarakat, baik yang mendukung maupun yang menentang. Secara umum, komentar yang dikumpulkan menunjukkan penggunaan singkatan, emoji, dan gaya bahasa yang beragam, serta campuran bahasa formal dan informal. Proses pengolahan komprehensif diperlukan sebelum analisis lebih lanjut karena sifat data media sosial yang dinamis dan heterogen.

Tabel 1. Daftar akun tiktok yang dilakukan scraping

Nama akun	Durasi	Waktu post	Jumlah Komentar
@kdmp.desa.dukuh	1 menit 37 detik	16-10-2025	618
@dylaa_fadylaa	28 detik	26-05-2025	114
@lingkardotnews	1 menit 13 detik	24-11-2025	729
@tasya_gallery	1 menit	27-05-2025	731
@amalia_desiana	47 detik	21-07-2025	274
@kemenkopri	2 menit 47 detik	29-10-2025	564
@gen.b_	1 menit 44 detik	18-11-2025	471
@lahaji_liwuku	1 menit 33 detik	04-09-2025	541
@sholehabs	21 detik	18-11-2025	63
@beritakilat1	59 detik	26-11-2025	249

Tabel 2. Data scraping dari tiktok

No	Content
1	berarti pedagang kecil toko2 dan bank2 akan gulung tikar semua semua pasti akan belanja semu ke koprasu warung dan toko mati pinjaman ke koprasu bank cuma tempat penyimpanan
2	percuma pa percuma 🙄🙄
3	Mantappppppo
4	mantab bangt
5	wow proyek tni ini cuan cuan soalnya di kelola tni

Tabel 3. Data scraping dari X

No	Content
1	pendamping desa siap kebersamaan musdessa koperasi merah putih Dari bumi merah putih* kita bangun desa bangun Indonesia desa terdepan untuk Indonesia* #MusdessaKDPM #TPPBerkelanjutan @prabowo @YandriSusanto @ArizaPatria https://t.co/htQ52vVoPL
2	Mari kita segera menindak lanjuti SE Menteri Desa PDT Np 8 Tahun 2025 melakukan percepatan Musyawarah Desa khusus untuk persetujuan dukungan pengembalian pinjaman Koperasi Desa Merah Putih https://t.co/16iaujHras
3	PENDAMPING DESA seluruh Nusantara mendampingi proses pembentukan dan penguatan Koperasi Desa Merah Putih #PendampingDesa #TPPKolaborasi https://t.co/LL3pF8UDAR
4	PENDAMPING DESA seluruh Nusantara mendampingi proses pembentukan dan penguatan Koperasi Desa Merah Putih @prabowo @ZUL_Hasan @YandriSusanto @ArizaPatria #PendampingDesa #TPPKolaborasi #TPPLampungTengah https://t.co/cL8pd4gFBL
5	PENDAMPING DESA seluruh Nusantara mendampingi proses pembentukan dan penguatan

No	Content
	Koperasi Desa Merah Putih @prabowo @YandriSusanto @ArizaPatria #PendampingDesa #TPPKolaborasi #tppindonesia #TPPSulbar https://t.co/h0NUrFxJcJ

Tabel 1 menunjukkan daftar akun TikTok yang disrape yang mengandung informasi seperti nama akun, durasi video, waktu posting dan jumlah komentar tentang Program Koperasi Desa Merah Putih. Tujuan penyajian tabel ini adalah untuk menunjukkan sumber data utama dan menjaga proses pengambilan data di platform TikTok menjadi jelas.

Tabel 2 menunjukkan hasil scraping komentar dari platform TikTok, yang mengandung teks komentar masyarakat tentang Program Koperasi Desa Merah Putih. Pada data pada tabel ini, menunjukkan berbagai opini publik yang muncul di TikTok. Dan pada tabel 3 menampilkan hasil scraping data dari platform X (Twitter) yang dikumpulkan menggunakan library Snsrape.

4.2. Distribusi Data Komentar

Tabel 4. Distribusi label tiktok

Label	Jumlah	Persentase
Positif	1.553	98,35%
Negatif	26	1,65%
Total	1.579	100%

Tabel 5. Distribusi label x

Label	Jumlah	Persentase
Positif	1.553	98,35%
Negatif	26	1,65%
Total	1.579	100%

Tabel 6. Distribusi label gabungan dua platform

Label	Jumlah	Persentase
Positif	3.944	83,05%
Negatif	805	16,95%
Total	4.749	100%

Hasil pelabelan data komentar menunjukkan bahwa distribusi sentimen pada masing-masing platform memiliki karakteristik yang berbeda. Tabel Distribusi Label TikTok menunjukkan bahwa dari 3.169 komentar, sentimen negatif mendominasi dengan 2.774, atau 87,54%, sedangkan sentimen positif hanya berjumlah 395, atau 12,46%. Ini menunjukkan bahwa sebagian besar pengguna TikTok menyampaikan kritik, keraguan, atau kekhawatiran terhadap Program Koperasi Desa Merah Putih, terutama terkait dampak program terhadap pemilik usaha kecil dan efeknya pada mereka.

Tabel Distribusi Label X (Twitter) menunjukkan kecenderungan sentimen yang sangat berbeda dibandingkan TikTok. Sentimen positif mendominasi secara signifikan dari 1.579 data komentar, dengan 1.553 komentar, atau 98,35%, dan sentimen negatif hanya 26 komentar, atau 1,65%. Ini menunjukkan bahwa platform X lebih banyak digunakan sebagai alat untuk memberikan dukungan dan menyebarkan

informasi resmi tentang Program Koperasi Desa Merah Putih, yang biasanya dikomunikasikan dengan bahasa formal dan bernada positif.

Selain itu, Tabel Distribusi Label Gabungan Dua Platform menunjukkan bahwa dari 4.749 komentar, sentimen positif tetap mendominasi dengan 3.944, atau 83,05%, dan sentimen negatif dengan 805, atau 16,95%. Ini terjadi meskipun data TikTok didominasi oleh sentimen negatif, tetapi distribusi sentimen secara keseluruhan tetap cenderung positif karena banyaknya sentimen positif di platform X.

Perbedaan dalam distribusi sentimen antara platform ini menunjukkan bahwa fitur dan tujuan masing-masing media digital sangat memengaruhi opini yang disampaikan. TikTok cenderung menjadi tempat bagi orang-orang untuk mengungkapkan pendapat emosional dan pribadi mereka, sementara X lebih banyak berfokus pada berbagi informasi dan mendukung kebijakan. Dalam proses klasifikasi sentimen, kondisi distribusi kelas yang tidak seimbang ini sangat diperhatikan.

4.3. PreProcessing Data

Tahap preprocessing dilakukan untuk meningkatkan kualitas data teks sebelum proses pemodelan. Data komentar hasil scraping dari platform X dan TikTok masih mengandung berbagai noise, seperti URL, mention, hashtag, emoji, tanda baca, angka, serta variasi penggunaan huruf besar dan kecil. Oleh karena itu, dilakukan beberapa tahapan preprocessing, yaitu case folding, cleaning, tokenizing, dan stopword removal.

Tabel 7. Data setelah preprocessing platform tiktok

Sebelum Preprocessing Data	Sesudah Preprocessing Data
berarti pedagang kecil toko2 dan bank2 akan gulung tikar semua semua pasti akan belanja semu ke koprasi warung dan toko mati pinjaman ke koprasi bank cuma tempat penyimpanan	pedagang kecil toko bank gulung tikar belanja koperasi warung toko mati pinjaman koperasi bank tempat penyimpanan
percuma pa percuma 🤔🤔	percuma
Mantappppppo	mantap
mantab bangt	mantap
wow proyek tni ini cuan cuan soalnya di kelola tni	proyek tni cuan dikelola tni

Tabel 8. Data setelah preprocessing paltform x

Sebelum Preprocessing Data	Sesudah Preprocessing Data
pendamping desa siap membersamai musdessus koperasi merah putih Dari bumi merah putih* kita bangun desa bangun Indonesia desa terdepan untuk Indonesia* #MusdessusKDPM #TPPBerkelanjutan @prabowo @YandriSusanto @ArizaPatria https://t.co/htQ52vVoPL	pendamping desa siap membersamai musdessus koperasi merah putih bumi merah putih bangun desa bangun indonesia desa terdepan indonesia

Sebelum Preprocessing Data	Sesudah Preprocessing Data
Mari kita segera menindak lanjuti SE Menteri Desa PDT Np 8 Tahun 2025 melakukan percepatan Musyawarah Desa khusus untuk persetujuan dukungan pengembalian pinjaman Koperasi Desa Merah Putih https://t.co/i6iauJHras	segera menindak lanjuti se menteri desa pdt np percepatan musyawarah desa khusus persetujuan dukungan pengembalian pinjaman koperasi desa merah putih
PENDAMPING DESA seluruh Nusantara mendampingi proses pembentukan dan penguatan Koperasi Desa Merah Putih #PendampingDesa #TPPKolaborasi https://t.co/LL3pF8UDAR	pendamping desa nusantara mendampingi proses pembentukan penguatan koperasi desa merah putih
PENDAMPING DESA seluruh Nusantara mendampingi proses pembentukan dan penguatan Koperasi Desa Merah Putih @prabowo @ZUL_Hasan @YandriSusanto @ArizaPatria #PendampingDesa #TPPKolaborasi #TPPLampungTengah https://t.co/cL8pd4gFBL	pendamping desa nusantara mendampingi proses pembentukan penguatan koperasi desa merah putih
PENDAMPING DESA seluruh Nusantara mendampingi proses pembentukan dan penguatan Koperasi Desa Merah Putih @prabowo @YandriSusanto @ArizaPatria #PendampingDesa #TPPKolaborasi #tppindonesia #TPPSulbar https://t.co/h0NUrFxCjC	pendamping desa nusantara mendampingi proses pembentukan penguatan koperasi desa merah putih

Pada tahap case folding, seluruh teks komentar diubah menjadi huruf kecil untuk menghindari perbedaan makna akibat variasi kapitalisasi. Selanjutnya, tahap cleaning dilakukan untuk menghapus URL, simbol, emoji, angka, serta karakter non-alfabet yang tidak memiliki kontribusi terhadap analisis sentimen. Tahap tokenizing memecah teks menjadi token berupa kata-kata penyusunnya, sehingga memudahkan proses analisis selanjutnya. Terakhir, stopword removal diterapkan untuk menghilangkan kata-kata umum seperti “dan”, “yang”, “di”, dan “ke” yang tidak memiliki makna sentimen secara signifikan.

Hasil dari tahapan preprocessing ini adalah teks komentar yang lebih bersih, terstruktur, dan relevan secara semantik. Data hasil preprocessing disimpan dalam kolom Clean_Comment dan digunakan sebagai input pada tahap ekstraksi fitur dan klasifikasi sentimen. Proses preprocessing terbukti mampu mengurangi kompleksitas data serta meningkatkan kualitas representasi fitur teks.

4.4. Hasil Pembobotan TF-IDF

Hasil ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) menghasilkan sekumpulan kata unik yang merepresentasikan keseluruhan korpus komentar masyarakat. Setiap kata diberi bobot numerik berdasarkan tingkat kemunculannya dalam satu dokumen serta tingkat kepentingannya terhadap seluruh dataset.

Sebagai contoh, fitur TF-IDF yang dihasilkan dalam penelitian ini antara lain: aamiin, abc, abdiyahdil, abdurrahman, abdy, abi, abimah, abis, abriawan, abs, absasal, abstrak, acara, acc, acceleratingrenewableenergy, aceh, achmadraihaan, ad, adadikompas, adakan.

Kata “aamiin” sering muncul pada komentar yang bersifat dukungan atau persetujuan, sehingga berkontribusi terhadap klasifikasi sentimen positif. Kata-kata seperti “abdurrahman”, “achmadraihaan”, dan “abriawan” merupakan nama tokoh atau akun yang sering disebut dalam diskusi terkait Program Koperasi Desa Merah Putih, terutama pada platform X, dan berfungsi sebagai penanda konteks percakapan.

Sementara itu, kata “acara”, “adakan”, dan “acc” umumnya muncul dalam komentar yang membahas kegiatan, pelaksanaan, atau persetujuan program, yang banyak ditemukan pada unggahan bersifat informatif. Kata “acceleratingrenewableenergy” dan “adadikompas” menunjukkan keterkaitan komentar dengan isu pembangunan, media, atau tagar tertentu, yang memperkaya konteks analisis sentimen.

Bobot TF-IDF yang diberikan pada kata-kata tersebut memungkinkan model untuk membedakan kata yang bersifat umum dengan kata yang memiliki nilai informatif tinggi dalam menentukan sentimen. Dengan representasi ini, setiap komentar tidak hanya dipandang sebagai teks mentah, tetapi sebagai vektor numerik yang mencerminkan kontribusi setiap kata terhadap kelas sentimen tertentu.

Hasil pembobotan TF-IDF ini selanjutnya digunakan sebagai input utama dalam proses klasifikasi menggunakan algoritma Support Vector Machine (SVM), sehingga model mampu mengenali pola sentimen positif dan negatif secara lebih akurat berdasarkan fitur-fitur kata yang signifikan.

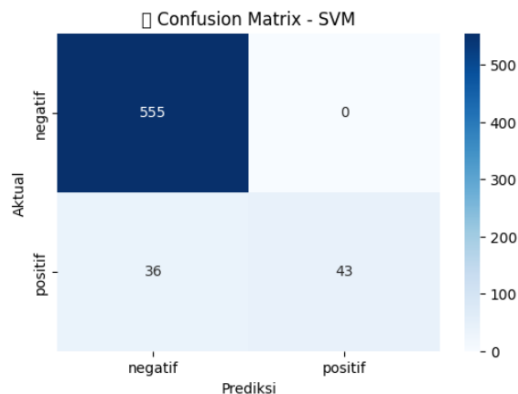
4.5. Hasil Model SVM di Platform Tiktok

Classification Report:				
	precision	recall	f1-score	support
negatif	0.94	1.00	0.97	555
positif	1.00	0.54	0.70	79
accuracy			0.94	634
macro avg	0.97	0.77	0.84	634
weighted avg	0.95	0.94	0.94	634

Gambar 2. Hasil model svm platform tiktok

Berbeda dengan platform X, hasil analisis sentimen pada platform TikTok menunjukkan bahwa sentimen negatif lebih dominan dibandingkan sentimen positif. Dari total 634 data uji, terdapat 555

komentar bersentimen negatif dan 79 komentar bersentimen positif. Hal ini menunjukkan bahwa pengguna TikTok cenderung mengekspresikan kritik, kekhawatiran, atau ketidaksetujuan terhadap Program Koperasi Desa Merah Putih secara lebih terbuka dan emosional.



Gambar 3. Confusion matrix svm platform tiktok

Hasil evaluasi model SVM pada platform TikTok menunjukkan akurasi sebesar 94,32%. Kelas sentimen negatif memiliki nilai precision sebesar 0,94, recall sebesar 1,00, dan F1-score sebesar 0,97, yang mengindikasikan bahwa model sangat efektif dalam mengidentifikasi komentar negatif.

Sebaliknya, kelas sentimen positif memperoleh precision sebesar 1,00, namun nilai recall hanya sebesar 0,54 dengan F1-score sebesar 0,70. Rendahnya recall pada kelas positif menunjukkan bahwa sebagian komentar positif masih diklasifikasikan sebagai negatif, yang disebabkan oleh dominasi data negatif dan variasi bahasa yang lebih bebas pada platform TikTok. Meskipun demikian, nilai weighted average F1-score sebesar 0,94 menunjukkan bahwa model tetap memiliki performa yang stabil dan dapat diandalkan.

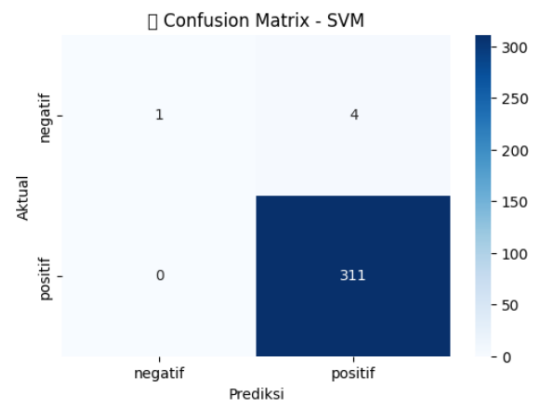
4.6. Hasil Model SVM di Platform Tiktok

Classification Report:				
	precision	recall	f1-score	support
negatif	1.00	0.20	0.33	5
positif	0.99	1.00	0.99	311
accuracy			0.99	316
macro avg	0.99	0.60	0.66	316
weighted avg	0.99	0.99	0.98	316

Gambar 4. Hasil svm platform x

Hasil analisis sentimen pada platform X (Twitter) menunjukkan bahwa sentimen positif sangat mendominasi dibandingkan sentimen negatif. Dari total data uji sebanyak 316 komentar, terdapat 311 komentar bersentimen positif dan hanya 5 komentar bersentimen negatif. Dominasi sentimen positif ini menunjukkan bahwa mayoritas pengguna X memberikan dukungan, persetujuan, atau

penyampaian informasi yang bernada positif terhadap Program Koperasi Desa Merah Putih.



Gambar 5. Confusion matrix svm platform x

Model klasifikasi sentimen menggunakan Support Vector Machine (SVM) berbasis TF-IDF pada platform X menghasilkan akurasi sebesar 98,73%. Berdasarkan classification report, kelas sentimen positif memperoleh nilai precision sebesar 0,99, recall sebesar 1,00, dan F1-score sebesar 0,99, yang menandakan bahwa hampir seluruh komentar positif dapat diklasifikasikan dengan sangat baik oleh model.

Sementara itu, kelas sentimen negatif memiliki precision sebesar 1,00, namun recall hanya sebesar 0,20 dengan F1-score sebesar 0,33. Nilai recall yang rendah ini dipengaruhi oleh jumlah data negatif yang sangat sedikit, sehingga model mengalami keterbatasan dalam mengenali seluruh komentar negatif secara optimal. Meskipun demikian, nilai weighted average F1-score sebesar 0,98 menunjukkan bahwa performa model secara keseluruhan pada platform X sangat baik.

4.7. Perbandingan Sentimen Positif dan Negatif Antar Platform

Berdasarkan hasil analisis sentimen yang telah dilakukan, dapat disimpulkan bahwa terdapat perbedaan pola opini masyarakat yang cukup signifikan antara platform X dan TikTok dalam menanggapi Program Koperasi Desa Merah Putih. Pada platform X, mayoritas komentar menunjukkan sentimen positif, yang mengindikasikan bahwa pengguna platform ini cenderung memberikan dukungan, apresiasi, serta menyebarkan informasi yang bersifat informatif dan normatif terkait program tersebut. Hal ini tidak terlepas dari karakteristik platform X yang sering digunakan sebagai media komunikasi resmi, penyampaian kebijakan, serta penyebaran narasi institusional oleh akun pemerintah maupun tokoh publik.

Sebaliknya, hasil analisis sentimen pada platform TikTok menunjukkan dominasi sentimen negatif, yang menggambarkan adanya kecenderungan pengguna untuk mengekspresikan kritik, keraguan, dan kekhawatiran terhadap implementasi Program Koperasi Desa Merah Putih. TikTok sebagai platform

berbasis konten audio-visual dan interaksi komentar yang lebih santai memungkinkan pengguna untuk menyampaikan opini secara lebih spontan, emosional, dan personal. Kondisi ini menyebabkan munculnya lebih banyak komentar yang bersifat evaluatif, kritis, serta mencerminkan pengalaman atau persepsi individu secara langsung.

Perbedaan kecenderungan sentimen antara kedua platform ini menunjukkan bahwa karakteristik media digital memiliki peran penting dalam membentuk pola opini publik. Faktor seperti tujuan penggunaan platform, gaya komunikasi pengguna, serta bentuk interaksi sosial yang tersedia turut memengaruhi bagaimana masyarakat menyampaikan pandangan mereka terhadap suatu kebijakan publik. Dengan demikian, analisis sentimen berbasis multi-platform memberikan gambaran yang lebih komprehensif mengenai persepsi masyarakat dibandingkan jika hanya menggunakan satu sumber data.

Meskipun memiliki karakteristik data dan distribusi sentimen yang berbeda, hasil evaluasi menunjukkan bahwa kombinasi metode Term Frequency–Inverse Document Frequency (TF-IDF) dan algoritma Support Vector Machine (SVM) mampu mengklasifikasikan sentimen positif dan negatif pada kedua platform dengan tingkat akurasi yang tinggi. Hal ini membuktikan bahwa pendekatan yang digunakan dalam penelitian ini efektif dan cukup robust untuk menangani data teks dari berbagai platform digital, sekaligus menegaskan potensi metode TF-IDF dan SVM dalam mendukung analisis sentimen terhadap kebijakan publik.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa analisis sentimen terhadap Program Koperasi Desa Merah Putih dengan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) dan algoritma Support Vector Machine (SVM) dapat menunjukkan persepsi masyarakat di berbagai platform digital, terutama Twitter dan TikTok.

Hasil analisis menunjukkan bahwa ada perbedaan signifikan dalam aspek sentimen antar platform. Disebabkan oleh fakta bahwa platform X banyak digunakan untuk komunikasi institusional, informasi resmi, dan narasi kebijakan, sebagian besar penggunaannya menunjukkan dukungan positif terhadap Program Koperasi Desa Merah Putih.

Sebaliknya, komentar negatif lebih banyak daripada komentar positif di TikTok. Komentar yang dibuat oleh pengguna TikTok biasanya mengandung kritik, keraguan, dan kekhawatiran tentang konsekuensi dari penerapan aplikasi, terutama yang berkaitan dengan usaha kecil dan ekonomi lokal. Perbedaan ini menunjukkan bahwa fitur media digital sangat memengaruhi cara orang berbicara.

Dari segi performa model, kombinasi TF-IDF dan SVM menunjukkan hasil yang sangat baik dalam mengklasifikasikan sentimen pada kedua platform. Akurasi model SVM adalah 94,32% pada platform

TikTok dan 98,73% pada platform X, masing-masing. Nilai akurasi, recall, dan F1-score secara keseluruhan menunjukkan bahwa metode ini efektif dan mampu menangani data teks yang sangat besar dengan distribusi kelas yang tidak seimbang. Dengan demikian, penelitian ini membuktikan bahwa pendekatan analisis sentimen berbasis TF-IDF dan SVM dapat digunakan secara andal untuk menganalisis opini publik terhadap kebijakan pemerintah berbasis data media sosial multi platform.

Penelitian selanjutnya disarankan untuk menggunakan metode pelabelan sentimen yang lebih akurat, seperti pelabelan manual atau pendekatan berbasis deep learning, guna meningkatkan kualitas klasifikasi sentimen. Selain itu, pengembangan penelitian dapat dilakukan dengan membandingkan algoritma klasifikasi lain, seperti Random Forest, Gradient Boosting, atau model berbasis neural network untuk memperoleh performa yang lebih optimal. Penambahan sumber data dari platform digital lain seperti YouTube, Facebook, atau Instagram juga diharapkan dapat memberikan gambaran opini publik yang lebih komprehensif. Bagi pemerintah dan pemangku kebijakan, hasil analisis sentimen ini dapat dijadikan sebagai bahan evaluasi dan pertimbangan dalam penyempurnaan implementasi Program Koperasi Desa Merah Putih agar lebih sesuai dengan kebutuhan dan aspirasi masyarakat.

DAFTAR PUSTAKA

- [1] R. Ridwan, D. Riswandi, and Y. S. Rachmanda, "DETEKSI DINI KECURANGAN PADA PROGRAM KOPERASI MERAH PUTIH: ANALISIS KELEMBAGAAN DESA DAN IMPLEMENTASI MODEL SIPAK," *J. Ekon. PERJUANGAN*, vol. 7, no. 1, pp. 41–53, 2025.
- [2] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, "Text mining in big data analytics," *Big Data Cogn. Comput.*, vol. 4, no. 1, p. 1, 2020.
- [3] V. A. Nugraha, "ANALISIS KONTEN KESEHATAN MENTAL PADA TIKTOK MENGGUNAKAN METODE SENTIMEN ANALISIS," *PREPOTIF J. Kesehat. Masy.*, vol. 9, no. 2 SE-Articles, pp. 3003–3008, Jun. 2025, doi: 10.31004/prepotif.v9i2.44638.
- [4] H. Zhou, "Research of Text Classification Based on TF-IDF and CNN-LSTM," *J. Phys. Conf. Ser.*, vol. 2171, no. 1, p. 12021, 2022, doi: 10.1088/1742-6596/2171/1/012021.
- [5] A. Roy and S. Chakraborty, "Support vector machine in structural reliability analysis: A review," *Reliab. Eng. Syst. Saf.*, vol. 233, p. 109126, 2023.
- [6] S. N. Cahyani and G. W. Saraswati, "Implementation of Support Vector Machine Method in Classifying School Library Books with Combination of TF-IDF and Word2Vec," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1555–1566, 2023.
- [7] R. Puspita and A. Widodo, "Perbandingan

- Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, pp. 646–654, 2021.
- [8] M. A. Khder, “Web scraping or web crawling: State of art, techniques, approaches and application,” *Int. J. Adv. Soft Comput. Its Appl.*, vol. 13, no. 3, 2021.
- [9] M. P. R. Putra and K. R. N. Wardani, “Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial,” *JUTIM (Jurnal Tek. Inform. Musirawas)*, vol. 5, no. 1, pp. 63–71, 2020.
- [10] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, “Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 422, 2021.
- [11] O. I. Gifari, M. Adha, I. R. Hendrawan, and F. F. S. Durrand, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022.
- [12] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, “Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM: Comparative Evaluation of SVM Kernels for Sentiment Classification in Fuel Price Increase Analysis,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023.
- [13] Q. Li and T. Li, “Research on the application of Naive Bayes and Support Vector Machine algorithm on exercises Classification,” in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 12071.
- [14] F. Fauzi, B. Andini, and M. Muslihudin, “SENTIMENT ANALYSIS OF TWITTER (X) USERS TOWARD THE KOPERASI DESA MERAH PUTIH PROGRAM USING SUPPORT VECTOR MACHINE AND NAIVE BAYES METHODS,” *J. TAM (Technology Accept. Model.)*, vol. 16, no. 1, pp. 58–64, 2025.
- [15] F. Novianti and K. R. N. Wardani, “Analisis Sentimen Masyarakat Terhadap Data Tweet Traveloka Selama Rapid Test Antigen Menggunakan Algoritma Naïve Bayes,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 3, pp. 922–933, 2023.
- [16] N. Sofi, T. Sulistyorini, and M. Nazaruddin, “Analisis Sentimen Masyarakat Pengguna Media Sosial Twitter Terhadap Motogp Mandalika Lombok Menggunakan Metode Bidirectional Encoder Representation From Transformers (BERT),” *J. Informasi, Sains dan Teknol.*, vol. 6, no. 1, pp. 120–130, 2023.
- [17] J. Foster *et al.*, “TikTok as a Source of Education and Misinformation in Lichen Sclerosus,” *J. Low. Genit. Tract Dis.*, pp. 10–1097, 2024.
- [18] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, “Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19),” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021.
- [19] H. H. MUZAKKIR, “IMPLEMENTASI ALGORITMA RULE BASED PADA SISTEM PENGAJUAN CUTI KARYAWAN BERBASIS WEB DI CV. SUMBER BAHAGIA PRINTING,” 2024, *UNIVERSITAS PGRI SEMARANG*.
- [20] W. L. Azzahra and E. Mailoa, “Analisis Sentimen Terhadap Rsud Salatiga Menggunakan Svm Dan Tf-Idf,” *J. Indones. Manaj. Inform. dan Komun.*, vol. 6, no. 1, pp. 478–489, 2025.
- [21] S. E. H. Yulianti, O. Soesanto, and Y. Sukmawaty, “Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit,” *J. Math. Theory Appl.*, pp. 21–26, 2022.
- [22] A. A. Ribeiro and M. Sachine, “On the optimal separating hyperplane for arbitrary sets: A generalization of the SVM formulation and a convex hull approach,” *Optimization*, vol. 71, no. 1, pp. 213–226, 2022.
- [23] Y. A. Prasetyo, E. Utami, and A. Yaqin, “Pengaruh Komposisi Split Data Terhadap Performa Akurasi Analisis Sentimen Algoritma Naïve Bayes dan SVM,” *J. homepage J. Electr. Eng. Comput.*, vol. 6, no. 2, 2024.