



Analisis Sentimen Penonton Terhadap Film Sore Istri dari Masa Depan pada Twitter Menggunakan Algoritma Support Vector Machine

Alfaza Putra Adjie Ariefiansyah*, Andri Firmansyah

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia

Email: ^{1,*}alfaza.312210512@mhs.pelitabangsa.ac.id, ²andrifirmansyah@pelitabangsa.ac.id

Email Penulis Korespondensi: alfaza.312210512@mhs.pelitabangsa.ac.id

Abstrak—Perkembangan media sosial telah mendorong masyarakat menyampaikan opini terhadap karya perfilman secara terbuka, cepat, dan real-time. Twitter atau X menjadi salah satu platform yang banyak digunakan untuk memberikan ulasan singkat terhadap film, sehingga menghasilkan data opini dalam jumlah besar yang sulit dianalisis secara manual. Penelitian ini bertujuan untuk menganalisis kecenderungan sentimen penonton terhadap film Sore: Istri dari Masa Depan berdasarkan data tweet, menerapkan algoritma Support Vector Machine dalam proses klasifikasi sentimen, serta mengevaluasi performa model yang digunakan. Data penelitian diperoleh melalui proses crawling tweet dengan kata kunci yang berkaitan dengan film Sore: Istri dari Masa Depan pada periode Juli 2025 sampai April 2026. Sebanyak 7.467 tweet dianalisis melalui text preprocessing, pelabelan sentimen, pembobotan fitur menggunakan Term Frequency-Inverse Document Frequency, dan klasifikasi menggunakan Support Vector Machine. Evaluasi model dilakukan menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score serta pembagian data training dan testing sebesar 80:20. Hasil penelitian menunjukkan bahwa 6.822 tweet tergolong sentimen positif dan 645 tweet tergolong sentimen negatif. Model memperoleh akurasi sebesar 90,75%, sehingga mampu mengklasifikasikan sentimen penonton secara cukup baik. Penelitian ini berkontribusi dalam menyediakan pendekatan komputasional untuk memetakan penerimaan publik terhadap film Indonesia melalui data media sosial, sekaligus memberikan dasar bagi pelaku industri perfilman dalam mengevaluasi respons penonton dan merancang strategi promosi berbasis data.

Kata Kunci: Analisis Sentimen; Twitter; Support Vector Machine; TF-IDF; Text Mining

Abstract—The development of social media has encouraged people to express their opinions about cinematic works openly, quickly, and in real time. Twitter, now known as X, has become one of the most widely used platforms for sharing brief film reviews, generating a large volume of opinion data that is difficult to analyze manually. This study aims to analyze audience sentiment toward the film Sore: Istri dari Masa Depan based on tweet data, apply the Support Vector Machine algorithm in the sentiment classification process, and evaluate the performance of the resulting model. The research data were collected through a tweet crawling process using keywords related to the film Sore: Istri dari Masa Depan from July 2025 to April 2026. A total of 7,467 tweets were analyzed through text preprocessing, sentiment labeling, feature weighting using Term Frequency-Inverse Document Frequency, and classification using Support Vector Machine. Model evaluation was conducted using a confusion matrix with accuracy, precision, recall, and F1-score metrics, using an 80:20 training and testing data split. The results showed that 6,822 tweets were classified as positive sentiment and 645 tweets as negative sentiment. The model achieved an accuracy of 90.75%, indicating that it was able to classify audience sentiment effectively. This study contributes by providing a computational approach to mapping public reception of Indonesian films through social media data, while also offering a basis for film industry practitioners to evaluate audience responses and develop data-driven promotional strategies.

Keywords: Sentiment Analysis; Twitter; Support Vector Machine; TF-IDF; Text Mining

1. PENDAHULUAN

Kemajuan teknologi informasi mengubah cara masyarakat menikmati, menilai, dan mendiskusikan karya hiburan, termasuk film. Pada masa sebelumnya, penilaian penonton lebih banyak muncul melalui percakapan langsung, media massa, atau forum ulasan tertentu. Saat ini, media sosial membuat opini penonton hadir secara cepat, spontan, terbuka, dan dapat diamati dalam skala besar. Kondisi tersebut menunjukkan bahwa penonton tidak lagi hanya menjadi penerima pesan film, tetapi juga menjadi bagian dari pembentuk wacana publik melalui komentar, kritik, pujian, rekomendasi, dan interpretasi pribadi. Twitter atau X menjadi salah satu media sosial yang relevan untuk kajian opini karena platform ini berbasis teks, menggunakan unggahan singkat, serta memungkinkan respons secara real-time terhadap topik populer. Data yang dihasilkan melalui Twitter dapat digunakan untuk memahami pola respons masyarakat terhadap suatu film, terutama ketika percakapan berlangsung dalam jumlah besar dan sulit dibaca satu per satu secara manual (Arum et al., 2023; Upa, 2024).

Film Sore: Istri dari Masa Depan merupakan film Indonesia yang menarik untuk dikaji karena menggabungkan unsur fiksi romantis, perjalanan waktu, hubungan antarmanusia, serta tema perubahan diri. Film ini memunculkan respons yang beragam dari pengguna media sosial, mulai dari apresiasi terhadap cerita, akting, dan pesan emosional hingga kritik terhadap alur, tempo, atau penyampaian konflik. Kajian terhadap film ini juga memiliki relevansi karena Sore: Istri dari Masa Depan tidak hanya hadir sebagai karya hiburan, tetapi juga sebagai teks budaya yang dibicarakan melalui ruang digital. Afriani et al. (2025) menyoroti film Sore dari sisi representasi destinasi wisata dan strategi promosi, sedangkan Setyaningsih & Rahmawati (2022) mengkaji mini seri Sore dari aspek tindak tutur ekspresif. Kedua penelitian tersebut menunjukkan bahwa karya Sore telah dikaji melalui pendekatan promosi, budaya, dan kebahasaan. Namun, penelitian tersebut belum menganalisis kecenderungan opini penonton secara komputasional berdasarkan data Twitter dalam jumlah besar. Oleh sebab itu, diperlukan kajian yang mampu memetakan penerimaan publik terhadap film tersebut melalui pendekatan analisis sentimen.



Analisis sentimen menjadi salah satu pendekatan yang dapat digunakan untuk mengolah data opini tidak terstruktur menjadi informasi yang lebih terukur. Analisis sentimen bertujuan mengidentifikasi kecenderungan opini dalam teks, misalnya positif atau negatif, dengan memanfaatkan teknik pemrosesan bahasa alami dan pembelajaran mesin. Data media sosial tidak selalu tersusun dengan baik karena mengandung singkatan, bahasa tidak baku, emoji, tanda baca berlebihan, tautan, mention, serta konteks yang sering kali implisit. Oleh sebab itu, tahapan text preprocessing seperti cleaning, case folding, tokenizing, normalisasi, stopword removal, dan stemming diperlukan agar teks lebih siap diproses oleh model klasifikasi (Puspitasari & Sutabri, 2024; Rifaldi & Fadlil, 2023). Setelah data dibersihkan, teks perlu direpresentasikan menjadi fitur numerik, salah satunya menggunakan Term Frequency-Inverse Document Frequency atau TF-IDF. Representasi tersebut membantu algoritma memberikan bobot lebih tinggi pada kata-kata yang memiliki peran penting dalam membedakan kecenderungan sentimen.

Algoritma Support Vector Machine atau SVM banyak digunakan dalam klasifikasi sentimen karena memiliki kemampuan memisahkan data ke dalam kelas tertentu melalui pencarian hyperplane terbaik. Sejumlah penelitian terdahulu menunjukkan bahwa SVM efektif digunakan untuk mengolah data teks media sosial. Khairudin et al. (2023) menerapkan SVM dalam analisis sentimen film di Twitter dan memperoleh akurasi sekitar 87,5% setelah proses penyesuaian parameter. Penelitian tersebut memiliki kesamaan dengan penelitian ini karena sama-sama menggunakan data opini film dari Twitter dan menerapkan SVM. Perbedaannya, penelitian ini secara khusus menganalisis film *Sore: Istri dari Masa Depan*, menggunakan sebanyak 7.467 tweet, serta mengevaluasi model melalui metrik accuracy, precision, recall, dan F1-score. Widodo et al. (2025) membandingkan Naive Bayes dan SVM pada analisis sentimen film *Jumbo*. Berbeda dengan penelitian tersebut yang berorientasi pada perbandingan algoritma, penelitian ini berfokus pada pengujian kemampuan SVM sebagai model utama untuk memetakan dua kelas sentimen penonton terhadap satu film Indonesia.

Ramdhania & Hidayat (2024) menunjukkan bahwa SVM menghasilkan akurasi yang lebih baik daripada Naive Bayes dalam analisis sentimen pengguna Twitter terhadap layanan Transjakarta. Kesamaannya terletak pada penggunaan data Twitter, TF-IDF, dan evaluasi model klasifikasi, sedangkan perbedaannya terdapat pada objek penelitian. Penelitian Ramdhania dan Hidayat berfokus pada layanan transportasi publik, sementara penelitian ini mengkaji penerimaan penonton terhadap karya perfilman. Safitri et al. (2023), Novitasari (2025), Hermawan & Jowensen (2023), Husada & Paramita (2021), serta Anggraini & Alita (2024) juga menerapkan SVM untuk memetakan opini publik pada objek musik, konser, produk, layanan penerbangan, isu publik, dan politik. Penelitian-penelitian tersebut menegaskan fleksibilitas SVM dalam mengolah berbagai objek opini di media sosial, tetapi belum secara khusus menjelaskan respons penonton terhadap film *Sore: Istri dari Masa Depan*.

Penelitian analisis sentimen terhadap film lokal Indonesia sebelumnya telah dilakukan pada objek yang berbeda. Sasongko & Hilda (2024) menganalisis sentimen terhadap film *Dirty Vote*, sedangkan Sinulingga & Sitorus (2024) mengkaji sentimen penonton terhadap film horor Indonesia. Kedua penelitian tersebut memiliki kesamaan dengan penelitian ini karena menggunakan opini masyarakat sebagai sumber analisis terhadap karya perfilman. Namun, perbedaannya terletak pada objek film, karakteristik dataset, periode pengambilan data, jumlah data, serta fokus evaluasi model. Film *Dirty Vote* memiliki hubungan kuat dengan isu politik, sedangkan film horor memiliki karakter respons penonton yang berbeda dengan film fiksi romantis. Penelitian ini memusatkan perhatian pada film *Sore: Istri dari Masa Depan* yang mengandung tema hubungan, perjalanan waktu, dan perubahan diri, sehingga pola sentimen yang muncul berpotensi menunjukkan karakteristik yang berbeda dari objek penelitian sebelumnya.

Berdasarkan perbandingan tersebut, celah penelitian terletak pada belum tersedianya kajian analisis sentimen yang secara khusus memetakan respons penonton terhadap film *Sore: Istri dari Masa Depan* menggunakan data Twitter dalam jumlah besar dan algoritma SVM. Penelitian ini menawarkan beberapa kontribusi. Pertama, penelitian ini memberikan kontribusi empiris berupa pemetaan kecenderungan sentimen penonton berdasarkan 7.467 tweet yang dikumpulkan pada periode Juli 2025 hingga April 2026. Kedua, penelitian ini memberikan kontribusi metodologis melalui penerapan rangkaian text mining yang mencakup proses crawling, seleksi data, text preprocessing, pelabelan, pembobotan TF-IDF, klasifikasi SVM, dan evaluasi menggunakan confusion matrix. Ketiga, penelitian ini memberikan kontribusi praktis dengan menyediakan gambaran berbasis data mengenai penerimaan publik terhadap film Indonesia, yang dapat dimanfaatkan oleh produser, pembuat film, dan pelaku industri kreatif untuk mengevaluasi respons penonton serta merancang strategi komunikasi dan promosi yang lebih tepat.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengetahui karakteristik sentimen penonton terhadap film *Sore: Istri dari Masa Depan* berdasarkan data Twitter, menerapkan algoritma Support Vector Machine dalam klasifikasi sentimen positif dan negatif, serta mengevaluasi performa model klasifikasi yang dibangun. Urgensi penelitian terletak pada kebutuhan mengolah opini publik yang tersedia dalam jumlah besar, bergerak cepat, dan tidak terstruktur agar dapat diubah menjadi informasi yang lebih sistematis dan bermakna. Hasil penelitian ini tidak dimaksudkan untuk menggantikan penilaian kritis dan interpretasi mendalam terhadap film, tetapi untuk memberikan gambaran kuantitatif mengenai kecenderungan respons penonton di media sosial sekaligus memperluas penerapan analisis sentimen pada kajian perfilman Indonesia. Secara khusus, kontribusi penelitian ini terletak pada penyediaan pemetaan empiris terhadap 7.467 tweet, pengujian terukur kombinasi TF-IDF dan SVM, serta penyediaan informasi praktis bagi industri film untuk mengevaluasi penerimaan penonton dan menyusun strategi komunikasi berbasis data.



2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis text mining untuk menganalisis sentimen penonton terhadap film *Sore: Istri dari Masa Depan* pada platform Twitter atau X. Objek penelitian berupa tweet berbahasa Indonesia yang membahas film tersebut pada periode Juli 2025 sampai April 2026. Pemilihan Twitter sebagai sumber data didasarkan pada karakteristiknya sebagai platform berbasis teks yang memungkinkan pengguna menyampaikan opini secara singkat, terbuka, dan real-time. Data yang digunakan diperoleh melalui proses crawling berdasarkan kata kunci yang berkaitan dengan film *Sore: Istri dari Masa Depan*. Hasil crawling kemudian disimpan dalam bentuk dataset untuk diproses lebih lanjut. Secara umum, penelitian ini mengikuti tahapan pengumpulan data, seleksi data, preprocessing, pelabelan sentimen, pembobotan fitur TF-IDF, klasifikasi menggunakan SVM, dan evaluasi model. Rangkaian tahapan tersebut disusun agar data media sosial yang tidak terstruktur dapat diubah menjadi informasi klasifikasi yang terukur (Fitri et al., 2024; Gautama et al., 2025).

Pada tahap pengumpulan data, tweet dikumpulkan menggunakan kata kunci yang relevan dengan film. Data yang diperoleh dari media sosial masih memiliki kemungkinan mengandung duplikasi, data kosong, unggahan tidak relevan, serta elemen yang tidak dibutuhkan dalam analisis. Oleh sebab itu, dilakukan seleksi data untuk memastikan bahwa tweet yang digunakan benar-benar berkaitan dengan topik penelitian. Setelah seleksi, data diproses melalui text preprocessing. Tahapan ini meliputi cleaning untuk menghapus URL, mention, karakter khusus, angka atau simbol yang tidak diperlukan, serta tanda baca berlebihan; case folding untuk menyeragamkan teks menjadi huruf kecil; tokenizing untuk memecah teks menjadi kata; normalisasi untuk memperbaiki kata tidak baku; stopword removal untuk menghapus kata umum yang tidak memberikan kontribusi besar terhadap sentimen; dan stemming untuk mengubah kata berimbuhan ke bentuk dasar. Tahap ini penting karena kualitas data teks berpengaruh langsung terhadap performa model klasifikasi (Puspitasari & Sutabri, 2024; Rifaldi & Fadlil, 2023).

Pelabelan sentimen dilakukan untuk mengelompokkan tweet ke dalam kategori positif dan negatif. Sentimen positif menunjukkan adanya ekspresi pujian, apresiasi, ketertarikan, atau respons menyenangkan terhadap film, sedangkan sentimen negatif menunjukkan kritik, kekecewaan, atau penilaian kurang baik. Setelah data memiliki label, proses pembobotan fitur dilakukan menggunakan TF-IDF. Metode ini mengubah kata menjadi bobot numerik berdasarkan frekuensi kemunculan kata pada dokumen dan tingkat kelangkaannya dalam keseluruhan korpus. Kata yang sering muncul pada dokumen tertentu tetapi tidak terlalu umum di seluruh dokumen memperoleh bobot lebih tinggi, sehingga dapat membantu model mengenali ciri sentimen. Data hasil TF-IDF kemudian digunakan sebagai masukan bagi algoritma SVM. Dalam penelitian ini, data dibagi menjadi data training dan testing dengan rasio 80:20 agar model dapat dilatih pada sebagian besar data dan diuji pada data yang belum pernah dipelajari.

Algoritma SVM digunakan untuk membangun model klasifikasi dua kelas, yaitu positif dan negatif. Model bekerja dengan mencari batas pemisah terbaik antara kelas berdasarkan fitur yang dihasilkan oleh TF-IDF. Evaluasi model dilakukan menggunakan confusion matrix dan metrik accuracy, precision, recall, serta F1-score. Accuracy digunakan untuk mengetahui proporsi prediksi benar dari keseluruhan data uji. Precision menunjukkan ketepatan prediksi pada suatu kelas, recall menunjukkan kemampuan model menemukan data yang benar pada kelas tersebut, dan F1-score menggambarkan keseimbangan antara precision dan recall. Penggunaan beberapa metrik diperlukan karena distribusi kelas sentimen tidak selalu seimbang, sehingga akurasi saja tidak cukup untuk menggambarkan kualitas model.

Untuk memperjelas proses penelitian secara sistematis, setiap tahapan dirangkum berdasarkan aktivitas yang dilakukan dan keluaran yang dihasilkan. Ringkasan ini menunjukkan hubungan antara proses pengumpulan data, pengolahan teks, pembentukan fitur, klasifikasi, dan evaluasi model. Tahapan penelitian tersebut disajikan pada Tabel 1.

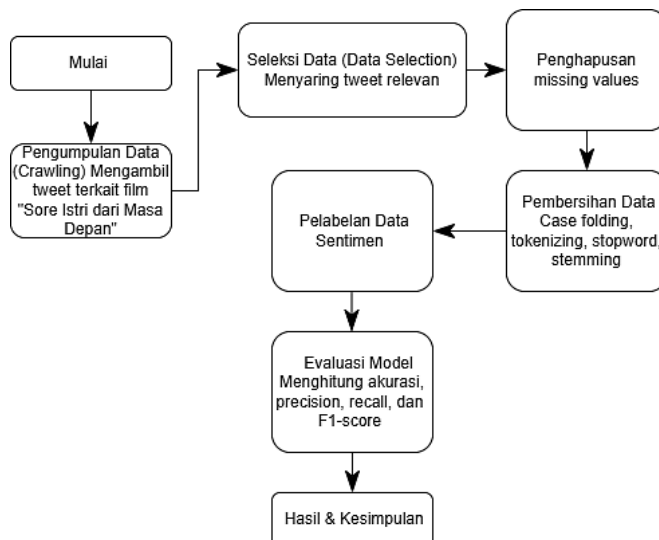
Tabel 1. Tahapan Penelitian

Tahapan	Proses	Hasil yang Diharapkan
Pengumpulan Data	Crawling tweet berdasarkan kata kunci film <i>Sore: Istri dari Masa Depan</i>	Dataset tweet mentah
Seleksi Data	Menghapus data kosong, duplikat, dan tidak relevan	Dataset relevan
Text Preprocessing	Cleaning, case folding, tokenizing, normalisasi, stopwords removal, dan stemming	Dataset bersih
Pelabelan Sentimen	Mengelompokkan tweet ke dalam sentimen positif dan negatif	Dataset berlabel
Pembobotan Fitur	Transformasi teks menggunakan TF-IDF	Fitur numerik
Klasifikasi	Penerapan algoritma Support Vector Machine	Model klasifikasi
Evaluasi	Pengujian menggunakan confusion matrix, accuracy, precision, recall, dan F1-score	Nilai performa model

Berdasarkan Tabel 1, penelitian dimulai dengan pengumpulan tweet yang berkaitan dengan film *Sore: Istri dari Masa Depan*. Data mentah kemudian diseleksi untuk menghapus tweet kosong, duplikat, dan tidak relevan. Data terpilih diproses melalui tahapan text preprocessing agar diperoleh teks yang lebih bersih dan seragam. Selanjutnya, data diberi label positif atau negatif, kemudian diubah menjadi fitur numerik menggunakan TF-IDF. Fitur tersebut digunakan

sebagai masukan bagi algoritma SVM untuk membangun model klasifikasi. Tahap terakhir adalah evaluasi model menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Dengan demikian, setiap tahap menghasilkan keluaran yang menjadi masukan bagi tahap berikutnya sampai diperoleh nilai performa model.

Selain disajikan dalam bentuk tabel, urutan pelaksanaan penelitian divisualisasikan dalam bentuk bagan alir agar hubungan antartahap dapat dipahami dengan lebih mudah. Bagan tersebut menggambarkan proses penelitian sejak pengumpulan data hingga penyusunan hasil dan kesimpulan, sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian

Gambar 1 menunjukkan bahwa penelitian diawali dengan proses crawling untuk mengambil tweet terkait film Sore: Istri dari Masa Depan. Data yang terkumpul selanjutnya memasuki tahap seleksi untuk menyaring tweet yang relevan. Data yang memiliki nilai kosong dihapus agar tidak mengganggu pengolahan, kemudian dilakukan pembersihan teks melalui case folding, tokenisasi, penghapusan stopword, normalisasi, dan stemming. Setelah teks menjadi lebih terstruktur, setiap tweet diberi label sentimen positif atau negatif. Data berlabel kemudian dibobotkan menggunakan TF-IDF dan diklasifikasikan menggunakan algoritma SVM. Model yang terbentuk dievaluasi berdasarkan nilai accuracy, precision, recall, dan F1-score. Hasil evaluasi tersebut digunakan sebagai dasar untuk menginterpretasikan kecenderungan sentimen penonton dan menyusun kesimpulan penelitian. Alur ini menunjukkan bahwa proses penelitian berlangsung secara berurutan dan setiap tahap memiliki fungsi dalam memastikan model klasifikasi dibangun menggunakan data yang relevan dan telah diproses dengan baik.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Hasil penelitian diperoleh dari proses analisis terhadap 7.467 tweet yang membahas film Sore: Istri dari Masa Depan pada platform Twitter. Data tersebut dikumpulkan melalui proses crawling berdasarkan kata kunci yang relevan dengan film, kemudian diseleksi dan dibersihkan agar dapat digunakan dalam klasifikasi. Pada data media sosial, preprocessing menjadi tahap yang sangat penting karena tweet sering memuat bahasa tidak baku, singkatan, tagar, mention, tautan, dan ekspresi emosional yang tidak selalu konsisten. Jika data tersebut tidak dibersihkan, model dapat menangkap noise sebagai fitur penting dan menghasilkan klasifikasi yang kurang tepat. Setelah proses preprocessing selesai, data diberi label sentimen positif dan negatif, kemudian diproses dengan TF-IDF untuk menghasilkan fitur numerik. Proses ini sejalan dengan penelitian Hermawan & Jowensen (2023), Khairudin et al. (2023), dan Safitri et al. (2023) yang menempatkan preprocessing dan pembobotan fitur sebagai tahapan penting dalam analisis sentimen berbasis Twitter.

Berdasarkan hasil pelabelan, sentimen penonton terhadap film Sore: Istri dari Masa Depan didominasi oleh sentimen positif. Dari total 7.467 tweet, sebanyak 6.822 tweet tergolong positif, sedangkan 645 tweet tergolong negatif. Distribusi tersebut menunjukkan bahwa respons pengguna Twitter terhadap film ini cenderung baik. Sentimen positif dapat muncul karena penonton menyukai cerita, pesan emosional, akting pemeran, visual, atau pengalaman menonton secara keseluruhan. Sementara itu, sentimen negatif menunjukkan adanya kritik terhadap aspek tertentu, misalnya alur cerita, tempo, atau ekspektasi penonton yang tidak terpenuhi. Distribusi sentimen secara rinci disajikan pada Tabel 2 dan divisualisasikan pada Gambar 2.

Tabel 2. Distribusi Sentimen Penonton

Kategori Sentimen	Jumlah Tweet	Persentase
Positif	6.822	91,36%

Berdasarkan Tabel 2, jumlah sentimen positif mencapai 6.822 tweet atau 91,36%, sedangkan sentimen negatif berjumlah 645 tweet atau 8,64%. Komposisi tersebut menunjukkan adanya ketimpangan kelas yang cukup besar dan menegaskan bahwa respons penonton terhadap film *Sore: Istri dari Masa Depan* didominasi oleh sentimen positif. Untuk memperjelas perbedaan proporsi antara sentimen positif dan negatif, distribusi data tersebut selanjutnya disajikan dalam bentuk diagram lingkaran pada Gambar 2.



Tabel 2 memperlihatkan bahwa sentimen positif mencapai 91,36%, sedangkan sentimen negatif hanya 8,64%. Perbedaan ini cukup besar dan mengindikasikan bahwa mayoritas perbincangan Twitter terhadap film *Sore: Istri dari Masa Depan* bernada apresiatif. Dalam konteks industri film, dominasi opini positif di media sosial dapat menjadi indikator awal bahwa film memperoleh penerimaan yang baik dari penonton digital. Respons positif yang disampaikan pengguna juga dapat berfungsi sebagai promosi organik karena unggahan di Twitter dapat menyebar melalui retweet, reply, dan interaksi antarpengguna. Walaupun demikian, jumlah sentimen negatif tetap perlu diperhatikan karena kritik penonton dapat menjadi masukan untuk memahami aspek yang dinilai kurang kuat. Interpretasi seperti ini sejalan dengan Sinulingga & Sitorus (2024), yang menempatkan data opini film sebagai sumber penting untuk membaca penerimaan publik terhadap karya sinema.

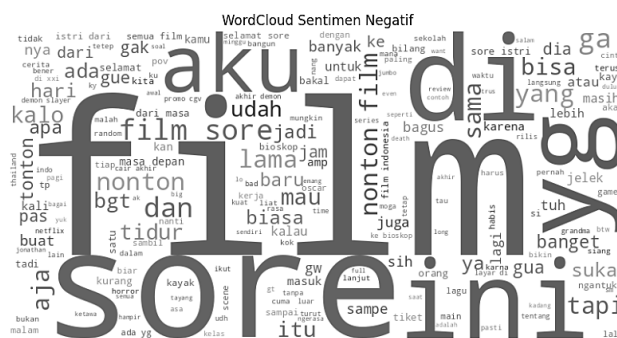
Visualisasi WordCloud digunakan untuk melihat kata-kata yang sering muncul pada setiap kelas sentimen. Pada kelas positif, kata-kata seperti “sore”, “istri”, “masa”, “depan”, “bagus”, “suka”, dan “banget” terlihat menonjol. Dominasi kata tersebut menunjukkan bahwa pembicaraan penonton berkisar pada judul film dan ekspresi pujian atau ketertarikan. Kata “suka” dan “bagus” menunjukkan bentuk apresiasi langsung, sedangkan kata “banget” sering berfungsi sebagai penguat emosi dalam bahasa percakapan media sosial. WordCloud sentimen positif ditampilkan pada Gambar 3. Pada kelas negatif, kata yang muncul antara lain “sore”, “film”, “kurang”, “lama”, dan “jelek”. Kata “kurang” dapat mengarah pada penilaian bahwa unsur tertentu belum memenuhi ekspektasi penonton, sedangkan kata “lama” dapat berkaitan dengan tempo cerita atau durasi yang dianggap kurang sesuai. WordCloud sentimen negatif ditampilkan pada Gambar 4.



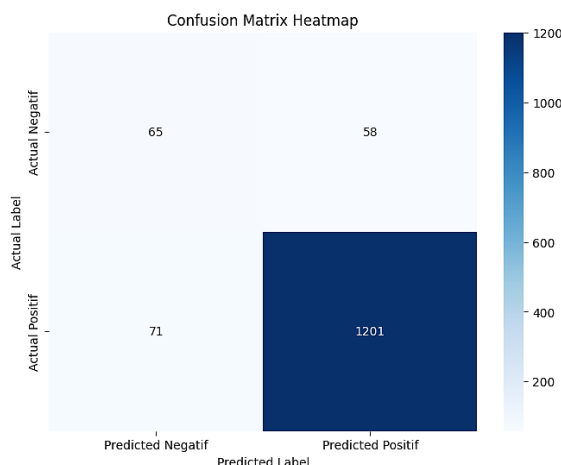
Gambar 3 memperlihatkan bahwa kata-kata yang berkaitan dengan judul film, seperti “film”, “sore”, “istri”, “masa”, dan “depan”, muncul secara dominan bersama kata-kata bernuansa apresiatif seperti “bagus”, “suka”, dan “banget”. Pola tersebut menunjukkan bahwa tweet positif tidak hanya menyebut objek film, tetapi juga memuat ekspresi ketertarikan dan penilaian yang baik dari penonton.

Selain sentimen positif, kata-kata yang dominan pada tweet negatif juga perlu ditampilkan untuk mengidentifikasi bentuk kritik atau ketidakpuasan penonton. Visualisasi kata pada kelas negatif disajikan pada Gambar 4. Gambar 4 menunjukkan kemunculan kata-kata seperti “kurang”, “lama”, dan “jelek” yang merepresentasikan penilaian negatif terhadap unsur tertentu dalam film. Walaupun frekuensinya lebih rendah daripada sentimen positif, kata-kata tersebut memberikan informasi penting mengenai aspek pengalaman menonton yang belum sepenuhnya memenuhi harapan sebagian pengguna Twitter.

Setelah data diproses dan diberi bobot menggunakan TF-IDF, model SVM dilatih dengan pembagian data training dan testing sebesar 80:20. Data uji yang digunakan berjumlah 1.395 tweet. Hasil confusion matrix menunjukkan bahwa model memprediksi 65 data negatif dengan benar, 58 data negatif sebagai positif, 71 data positif sebagai negatif, dan 1.201 data positif dengan benar. Dengan demikian, jumlah prediksi benar adalah 1.266 data dari total 1.395 data uji. Hasil tersebut menghasilkan akurasi sebesar 90,75%. Confusion matrix model ditunjukkan pada Gambar 5, sedangkan ringkasan metrik evaluasi disajikan pada Tabel 3.



Gambar 4. WordCloud Sentimen Negatif



Gambar 5. Confusion Matrix Heatmap Model SVM

Berdasarkan Gambar 5, model berhasil mengklasifikasikan 65 tweet negatif dan 1.201 tweet positif secara benar. Namun, masih terdapat 58 tweet negatif yang diprediksi sebagai positif serta 71 tweet positif yang diprediksi sebagai negatif. Hasil ini menunjukkan bahwa model lebih kuat dalam mengenali kelas positif dibandingkan kelas negatif, yang berkaitan dengan dominasi data positif dalam dataset.

Untuk memberikan penilaian yang lebih lengkap terhadap hasil pada confusion matrix, performa model dianalisis menggunakan precision, recall, F1-score, accuracy, macro average, dan weighted average. Ringkasan hasil evaluasi tersebut disajikan pada Tabel 3.

Tabel 3. Hasil Evaluasi Model SVM

Kelas/Metrik	Precision	Recall	F1-score	Support
Negatif	47,79%	52,85%	50,19%	123
Positif	95,39%	94,42%	94,89%	1.272
Accuracy	90,75%	90,75%	90,75%	1.395
Macro average	71,59%	73,63%	72,54%	1.395
Weighted average	91,20%	90,75%	90,96%	1.395

Tabel 3 menunjukkan bahwa model memiliki performa sangat baik dalam mengenali sentimen positif, tetapi performanya lebih rendah pada sentimen negatif. Nilai precision kelas positif sebesar 95,39% berarti sebagian besar tweet yang diprediksi positif memang benar positif. Nilai recall positif sebesar 94,42% menunjukkan bahwa model



mampu menemukan sebagian besar tweet positif yang ada pada data uji. F1-score positif sebesar 94,89% memperlihatkan keseimbangan yang kuat antara precision dan recall untuk kelas mayoritas. Sebaliknya, precision negatif sebesar 47,79%, recall negatif sebesar 52,85%, dan F1-score negatif sebesar 50,19% menunjukkan bahwa model masih mengalami kesulitan dalam mengenali sentimen negatif. Hal ini dapat dipengaruhi oleh distribusi data yang tidak seimbang karena jumlah sentimen positif jauh lebih banyak dibandingkan sentimen negatif.

Nilai akurasi 90,75% menunjukkan bahwa model SVM mampu mengklasifikasikan sebagian besar data uji secara benar. Namun, pembacaan performa model tidak boleh hanya berhenti pada akurasi. Pada dataset yang tidak seimbang, akurasi tinggi dapat dipengaruhi oleh dominasi kelas mayoritas. Dalam penelitian ini, kelas positif jauh lebih banyak daripada kelas negatif, sehingga model lebih mudah mempelajari pola positif. Kondisi tersebut menjelaskan mengapa precision, recall, dan F1-score kelas positif lebih tinggi dibandingkan kelas negatif. Temuan ini memperkuat pentingnya penggunaan metrik evaluasi yang beragam, sebagaimana ditegaskan oleh Nurhidayat & Dewi (2023), bahwa evaluasi lintas metrik diperlukan agar performa model klasifikasi dapat dibaca secara lebih proporsional.

Jika dibandingkan dengan penelitian sebelumnya, hasil penelitian ini menunjukkan performa yang kompetitif. Khairudin et al. (2023) memperoleh akurasi sekitar 87,5% pada analisis sentimen film di Twitter menggunakan SVM, sedangkan penelitian ini memperoleh akurasi 90,75%. Perbedaan tersebut dapat dipengaruhi oleh jumlah data, karakteristik topik, kualitas preprocessing, skema pelabelan, dan distribusi kelas. Widodo et al. (2025) pada film Jumbo juga menunjukkan bahwa SVM dapat digunakan untuk menganalisis sentimen penonton film, meskipun performa setiap model sangat bergantung pada karakteristik dataset. Sementara itu, Ramdhania & Hidayat (2024) memperlihatkan bahwa SVM mampu mengungguli Naive Bayes pada analisis sentimen pengguna Twitter terhadap Transjakarta. Perbandingan ini menegaskan bahwa SVM merupakan algoritma yang relevan untuk klasifikasi teks, terutama ketika dipadukan dengan preprocessing dan TF-IDF.

Dalam konteks objek penelitian, dominasi sentimen positif terhadap film *Sore: Istri dari Masa Depan* menunjukkan bahwa film ini memperoleh penerimaan yang baik dari pengguna Twitter. Hasil tersebut dapat dibaca sebagai bentuk keterlibatan emosional penonton terhadap cerita dan karakter. Kata-kata dominan pada WordCloud positif memperlihatkan bahwa pengguna Twitter banyak menyebut unsur judul film serta ekspresi apresiasi seperti suka dan bagus. Sementara itu, sentimen negatif yang jumlahnya lebih kecil tetap memberikan informasi penting mengenai kritik penonton. Kata-kata seperti kurang, lama, dan jelek menunjukkan adanya bagian dari pengalaman menonton yang dianggap belum sesuai harapan. Dengan demikian, hasil analisis sentimen tidak hanya memperlihatkan tingkat penerimaan, tetapi juga membantu mengidentifikasi sisi apresiasi dan kritik dalam percakapan digital.

3.2 Pembahasan

Temuan penelitian menunjukkan bahwa SVM mencapai akurasi 90,75% dalam mengklasifikasikan sentimen penonton film *Sore: Istri dari Masa Depan*. Nilai ini lebih tinggi daripada akurasi 87,5% yang dilaporkan Khairudin et al. (2023) pada analisis sentimen film di Twitter dan lebih tinggi daripada akurasi SVM 75,87% pada penelitian film Jumbo oleh Widodo et al. (2025). Perbedaan tersebut dapat dipengaruhi oleh jumlah data, karakteristik kosakata, teknik pelabelan, kualitas preprocessing, pembobotan fitur, serta distribusi kelas. Dataset penelitian ini berjumlah 7.467 tweet, sehingga model memperoleh variasi data yang lebih luas untuk mempelajari pola sentimen dibandingkan beberapa penelitian film terdahulu.

Jika dibandingkan dengan objek nonfilm, performa penelitian ini juga tergolong kompetitif. Ramdhania et al. (2024) memperoleh akurasi SVM 84,95% pada sentimen layanan Transjakarta, sedangkan Sinulingga dan Sitorus (2024) memperoleh akurasi 82,51% pada sentimen film horor Indonesia. Di sisi lain, akurasi penelitian ini sedikit lebih rendah daripada akurasi pengujian SVM 91,78% yang diperoleh Sari dan Widayat (2025) pada sentimen PON 2024. Perbandingan tersebut menunjukkan bahwa SVM cenderung konsisten pada data teks media sosial, tetapi performanya tetap dipengaruhi oleh konteks objek, skema pelabelan, representasi fitur, dan komposisi kelas pada setiap dataset.

Walaupun akurasi keseluruhan tinggi, F1-score kelas negatif sebesar 50,19% masih jauh di bawah F1-score kelas positif sebesar 94,89%. Kondisi ini menegaskan bahwa ketidakseimbangan data—6.822 tweet positif berbanding 645 tweet negatif—membuat model lebih kuat mengenali kelas mayoritas. Temuan tersebut sejalan dengan penelitian Sari dan Widayat (2025), yang juga menghadapi dominasi sentimen positif, serta menunjukkan bahwa evaluasi tidak cukup hanya mengandalkan akurasi. Oleh karena itu, penelitian selanjutnya perlu menerapkan teknik balancing seperti SMOTE, class weighting, atau undersampling untuk meningkatkan kemampuan model dalam mengenali sentimen negatif.

Secara metodologis, penelitian ini memperlihatkan bahwa kombinasi text preprocessing, TF-IDF, dan SVM dapat digunakan untuk memetakan opini penonton film lokal pada media sosial. Hasil ini memperluas konteks penelitian SVM yang sebelumnya banyak digunakan pada layanan publik, produk, isu politik, atau topik hiburan lain. menggunakan SVM untuk menganalisis sentimen konser, Novitasari (2025) menggunakannya pada opini produk skincare, sedangkan Anggraini & Alita (2024) menggunakannya pada sentimen politik. Penelitian ini menempatkan film lokal sebagai objek kajian dan menunjukkan bahwa data Twitter dapat memberikan gambaran respons publik terhadap karya perfilman Indonesia. Oleh karena itu, penelitian ini memiliki kontribusi pada bidang text mining sekaligus dapat menjadi rujukan praktis bagi industri film dalam memahami respons penonton berbasis data digital.

Meskipun demikian, hasil penelitian perlu dibaca dengan mempertimbangkan keterbatasan. Pertama, data hanya berasal dari Twitter sehingga belum mewakili seluruh opini penonton pada platform lain seperti Instagram, YouTube, TikTok, atau situs ulasan film. Kedua, penelitian hanya menggunakan dua kelas sentimen, yaitu positif dan negatif,



sehingga opini netral atau informatif belum dianalisis secara khusus. Ketiga, distribusi kelas yang tidak seimbang memengaruhi kemampuan model dalam mengenali sentimen negatif. Oleh karena itu, penelitian berikutnya dapat menggunakan data lintas platform, menambahkan kelas netral, menerapkan teknik balancing data, dan membandingkan SVM dengan algoritma lain seperti Naive Bayes, Random Forest, K-Nearest Neighbor, atau model berbasis deep learning. Rekomendasi tersebut selaras dengan penelitian Fachriza (2023), Sari & Widayat (2025), serta Septini & Susanto (2025) yang memperlihatkan bahwa variasi algoritma dan strategi evaluasi dapat memengaruhi hasil analisis sentimen pada data media sosial.

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa analisis sentimen terhadap 7.467 tweet mengenai film *Sore: Istri dari Masa Depan* pada Twitter mampu menjawab seluruh rumusan masalah penelitian. Pertama, karakteristik sentimen penonton menunjukkan kecenderungan yang didominasi oleh sentimen positif. Sebanyak 6.822 tweet atau 91,36% termasuk dalam kategori positif, sedangkan 645 tweet atau 8,64% termasuk dalam kategori negatif. Dominasi tersebut menunjukkan bahwa film *Sore: Istri dari Masa Depan* memperoleh penerimaan yang baik dari pengguna Twitter, terutama melalui ungkapan apresiasi, ketertarikan, dan respons emosional terhadap cerita serta pengalaman menonton, meskipun masih ditemukan kritik terhadap alur, tempo, dan beberapa aspek film. Kedua, algoritma Support Vector Machine berhasil diterapkan untuk mengklasifikasikan sentimen positif dan negatif melalui tahapan crawling data, seleksi data, penghapusan missing values, text preprocessing, pelabelan sentimen, pembobotan fitur menggunakan TF-IDF, pembagian data training dan testing dengan rasio 80:20, serta evaluasi menggunakan confusion matrix. Ketiga, berdasarkan pengujian terhadap 1.395 data uji, model memperoleh akurasi sebesar 90,75%, yang menunjukkan bahwa SVM mampu mengklasifikasikan sebagian besar tweet dengan benar. Model memiliki performa sangat baik dalam mengenali sentimen positif, tetapi masih mengalami kesulitan dalam mengenali sentimen negatif karena jumlah data positif jauh lebih besar daripada data negatif. Dengan demikian, kombinasi text preprocessing, TF-IDF, dan SVM efektif digunakan untuk memetakan kecenderungan opini penonton terhadap film Indonesia pada media sosial. Kontribusi penelitian ini mencakup bukti empiris mengenai dominasi penerimaan positif terhadap film Indonesia, validasi kinerja kombinasi TF-IDF dan SVM pada dataset tweet berskala besar, serta penyediaan informasi berbasis data yang dapat dimanfaatkan pelaku industri perfilman untuk mengevaluasi respons penonton dan merancang strategi promosi. Keterbatasan penelitian ini terletak pada penggunaan data yang hanya berasal dari Twitter, klasifikasi yang terbatas pada dua kelas sentimen, dan distribusi kelas yang tidak seimbang. Penelitian selanjutnya disarankan menggunakan data dari beberapa platform, menambahkan kategori netral, menerapkan teknik penyeimbangan data, serta membandingkan SVM dengan algoritma lain agar hasil analisis menjadi lebih komprehensif dan representatif.

REFERENCES

- Afriani, A. L., Nathanael, G. K., & Wardani, A. (2025). Soft-selling pariwisata dalam film *Sore: Istri dari Masa Depan* (Studi kasus representasi destinasi wisata dan strategi promosi). *MUKASI: Jurnal Ilmu Komunikasi*, 4(3), 1124–1138. <https://doi.org/10.54259/mukasi.v4i3.5410>
- Anggraini, J., & Alita, D. (2024). Implementasi metode SVM pada sentimen analisis terhadap pemilihan presiden (Pilpres) 2024 di Twitter. *Jurnal Penelitian Ilmu Dan Teknologi Komputer*, 9(2), 102–111. <https://doi.org/10.30591/jpit.v9i2.6560>
- Arum, D. S., Butsianto, S., & Astuti, R. (2023). Analisis sentimen masyarakat Indonesia terhadap Sea Games 2023 di Twitter dengan metode Naïve Bayes. *Journal of Information System, Applied, Management, Accounting and Research*, 7(3), 728–738. <https://doi.org/10.52362/jisamar.v7i3.1150>
- Fachriza, M. (2023). Analisis sentimen kalimat depresi pada pengguna Twitter dengan Naive Bayes, Support Vector Machine, dan Random Forest. *Jurnal Informatika Dan Komputer*, 49–58. <https://doi.org/10.24269/jkt.v7i2.2218>
- Fitri, S. D., Lestari, D., Bintana, R. R., & Aryani, R. (2024). Implementasi model Support Vector Machine dalam analisa sentimen masyarakat mengenai kebijakan penerapan aplikasi MyPertamina. *Bridge: Jurnal Publikasi Sistem Informasi Dan Telekomunikasi*, 2, 176–193. <https://doi.org/10.62951/bridge.v2i3.180>
- Gautama, P. J. J. S., Wibowo, A., & Siang, J. J. (2025). Analisis sentimen pengguna terhadap akun X/Twitter resmi “DANA” dengan algoritma IndoBERT. *INTECOMS: Journal of Information Technology and Computer Science*, 8(2), 549–559. <https://doi.org/10.31539/intecom.v8i2.14477>
- Hermawan, A., & Jowensen, I. (2023). Implementasi text-mining untuk analisis sentimen pada Twitter dengan algoritma Support Vector Machine. *Jurnal Informatika*, 12(1), 129–137. <https://doi.org/10.23887/jstundiksha.v12i1.52358>
- Husada, H. C., & Paramita, A. S. (2021). Analisis sentimen pada maskapai penerbangan di platform Twitter menggunakan algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18–26. <https://doi.org/10.34148/teknika.v10i1.311>
- Khairudin, M., Sukendar, A., & Somantri, A. (2023). Analisis sentimen film di Twitter menggunakan metode Support Vector Machine. *Jurnal Informatika*, 5(1), 97–102. <https://doi.org/10.59811/sandi.v5i1.47>
- Novitasari, D. T. (2025). Analisis sentimen pengguna Twitter terhadap skincare dengan metode Support Vector Machine (SVM). *Jurnal Sistem Informasi*, 19(2), 325–332. <https://doi.org/10.33480/inti.v19i2.6297>



- Nurhidayat, R., & Dewi, K. E. (2023). Penerapan algoritma K-Nearest Neighbor dan fitur ekstraksi N-Gram dalam analisis sentimen berbasis aspek. *KOMPUTA: Jurnal Ilmiah Komputer Dan Informatika*, 12(1), 91–100. <https://doi.org/10.34010/komputa.v12i1.9458>
- Puspitasari, D., & Sutabri, T. (2024). Analisis sentimen berdasarkan pada Twitter (X) terhadap layanan IndiHome menggunakan algoritma Support Vector Machine (SVM). *Jurnal Manajemen Informatika Dan Teknologi*, 3(2), 58–71. <https://doi.org/10.55123/jumintal.v3i2.4449>
- Ramdhania, K. F., & Hidayat, D. F. (2024). Implementasi metode Naïve Bayes dan Support Vector Machine (SVM) untuk menganalisis sentimen pengguna Twitter terhadap Transjakarta. *Jurnal Teknologi Informasi*, 9(1), 1–14. <https://doi.org/10.26594/jmpm.v9i1.4494>
- Rifaldi, H. D., & Fadlil, A. (2023). Teknik preprocessing pada text mining menggunakan data tweet “Mental Health.” *Jurnal Teknologi Informasi*, 3(2), 161–171. <https://doi.org/10.51454/decode.v3i2.131>
- Safitri, T., Umaidah, Y., & Maulana, I. (2023). Analisis sentimen pengguna Twitter terhadap BTS menggunakan algoritma Support Vector Machine. *Jurnal Informatika*, 7(1), 34–41. <https://doi.org/10.30871/jaic.v7i1.5039>
- Sari, F. D., & Widayat, W. (2025). Analisis sentimen masyarakat terhadap PON 2024 Aceh dan Sumatra Utara di Twitter menggunakan algoritma SVM dan KNN. *Jurnal Sistem Informasi*, 22(2), 251–265. <https://doi.org/10.24246/aiti.v22i2.251-265>
- Sasongko, K. H., & Hilda, A. M. (2024). Analisis sentimen terhadap film Dirty Vote pada media sosial X dan YouTube dengan algoritma Naive Bayes dan SVM. *Journal of Computer System and Informatics*, 6(1), 73–85. <https://doi.org/10.47065/josyc.v6i1.6150>
- Septini, E. A., & Susanto. (2025). Analisis sentimen masyarakat di Twitter mengenai OpenAI ChatGPT menggunakan metode Support Vector Machine (SVM). *Bulletin of Computer Science Research*, 5(2), 138–149. <https://doi.org/10.47065/bulletincsr.v5i2.475>
- Setyaningsih, M., & Rahmawati, L. E. (2022). Bentuk tindak tutur ekspresif dalam mini seri Sore: Istri dari Masa Depan karya Yandy Laurens. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 12, 85–98. <https://doi.org/10.23887/jjpbs.v12i1.43933>
- Sinulingga, J. E. B., & Sitorus, H. C. K. (2024). Analisis sentimen masyarakat terhadap film horor Indonesia menggunakan metode SVM dan TF-IDF. *Jurnal Manajemen Informatika*, 14, 42–53. <https://doi.org/10.34010/jamika.v14i1.11946>
- Upa, N. F. (2024). Analisis sentimen pengguna Twitter menggunakan metode SVM-KNN. *Nusantara Hasana Journal*, 3(12), 57–64. <https://doi.org/10.59003/nhj.v3i12.1125>
- Widodo, T. R., Fajri, I. N., & Sari, B. W. (2025). Sentiment analysis of the film Jumbo on Twitter using the Naive Bayes method and Support Vector Machine (SVM) with a text mining approach. *Jurnal Informatika*, 9(5), 2861–2868. <https://doi.org/10.30871/jaic.v9i5.10557>