ANALISIS SENTIMEN FILM AGAK LAEN MENYALA PANTIKU MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Nadhilah Al Adawiyah, Nadia Ramadhani, Salsa Yurinka, Winda Kurnia Sari, Ken Ditha Tania
Program Studi Sistem Informasi, Universitas Sriwijaya
Jl. Raya Palembang-Prabumulih No. KM 32 Inderalaya, Kabupaten Ogan Ilir, Sumatera Selatan (30662)
09031282328061@student.unsri.ac.id

ABSTRAK

Media sosial merupakan salah satu sarana bagi publik untuk menyampaikan opini dan ulasan terhadap berbagai topik, termasuk film, yang dapat memengaruhi persepsi publik secara luas. Film "Agak Laen: Menyala Pantiku!" merupakan salah satu film yang ramai diperbincangkan di platform X sehingga menghasilkan beragam tanggapan dari pengguna. Banyaknya data opini yang tersedia dalam bentuk teks tidak terstruktur menyebabkan proses analisis secara manual menjadi kurang efektif. Oleh karena itu, penelitian ini bertujuan untuk mengklasifikaskan sentimen opini pengguna platform X terhadap film "Agak Laen: Menyala Pantiku!" ke dalam tiga kategori positif, negatif, dan netral. Data penelitian dikumpulkan melalui proses *crawling* dari platform X, kemudian diproses melalui tahap *preprocessing* teks serta pembobotan fitur menggunakan metode TF-IDF. Proses klasifikasi kemudian diterapkan menggunakan algoritma *Support Vector Machine* (SVM). Berdasarkan hasil pengujian, model SVM mampu mengklasifikasikan sentimen dengan tingkat akurasi sebesar 80%. Distribusi sentimen menunjukkan bahwa sentimen positif memiliki jumlah paling tinggi, diikuti oleh sentimen netral, sedangkan sentimen negatif relatif lebih sedikit. Temuan ini mengindikasikan bahwa sebagian besar tanggapan publik terhadap film tersebut bersifat positif, terutama terkait unsur komedi yang menjadi daya tarik utama film.

Kata kunci : *Analisis Sentimen, Text Mining, Support Vector Machine, TF-IDF, Film*

1. PENDAHULUAN

Industri film saat ini semakin dipengaruhi oleh respons dan ulasan masyarakat yang disampaikan melalui media sosial. Diskusi daring yang terjadi di berbagai platform tidak hanya menjadi sarana berbagi opini, tetapi juga turut membentuk persepsi publik terhadap suatu karya film. Salah satu platform yang aktif digunakan untuk menyampaikan opini, kritik, maupun apresiasi adalah platform X. Melalui unggahan singkat yang tersebar secara luas, pengguna dapat membentuk arus percakapan yang memengaruhi cara masyarakat memandang sebuah film.

Salah satu film yang menarik perhatian publik adalah "Agak Laen: Menyala Pantiku!" yang tercatat sebagai film terlaris di Indonesia dengan jumlah penonton mencapai 11.012.445 orang berdasarkan data [1]. Tingginya jumlah penonton menunjukkan besarnya minat masyarakat terhadap film tersebut. Namun, popularitas film tidak hanya dapat diukur melalui jumlah penonton atau pendapatan *box office*, tetapi juga dapat dilihat dari bagaimana persepsi dan respons publik yang berkembang di media sosial [2]. Opini masyarakat yang tercermin melalui sentimen positif, negatif, maupun netral dapat memberikan gambaran mengenai penerimaan publik terhadap sebuah film serta berpotensi memengaruhi minat masyarakat lainnya untuk menonton di bioskop [3].

Di sisi lain, opini yang disampaikan pengguna media sosial umumnya berbentuk teks yang tidak terstruktur dan tersebar dalam jumlah besar, sehingga analisis secara manual sulit dilakukan. Oleh karena itu, diperlukan pendekatan analitis yang mampu mengolah dan mengelompokkan opini publik secara otomatis berdasarkan kecenderungan sentimennya. Analisis sentimen menjadi salah satu pendekatan dalam bidang pengolahan bahasa alami (*Natural Language Processing*) yang dapat mengidentifikasi dan mengklasifikasikan opini masyarakat ke dalam kategori sentimen tertentu [4].

Sejumlah penelitian sebelumnya telah menerapkan analisis sentimen untuk memahami respons publik terhadap film Indonesia. Penelitian terhadap film "KKN di Desa Penari" berhasil mencapai akurasi sebesar 60.7073% dengan memanfaatkan model Naïve Bayes [5]. Teknik serupa juga diimplementasikan oleh peneliti lain untuk menganalisis sentimen film animasi "*Jumbo"* di platform TikTok, dengan perolehan akurasi sebesar 63,40% [6]. Selain itu, penelitian yang dilakukan oleh [7] mengenai analisis sentimen film "Agak Laen" memperoleh tingkat akurasi sebesar 75.73% dalam mengklasifikasikan opini masyarakat menggunakan algoritma Naïve Bayes.

Meskipun Naïve Bayes cukup efektif dalam berbagai penelitian analisis sentimen, metode tersebut masih memiliki keterbatasan dalam menangani data teks yang kompleks dan berdimensi tinggi. Penelitian [7] juga menyarankan bahwa optimalisasi tahapan pra-pemrosesan data serta penggunaan algoritma alternatif seperti *Support Vector Machine* (SVM) atau *Random Forest* berpotensi meningkatkan performa klasifikasi secara signifikan. Namun demikian, penelitian yang mengkaji penggunaan algoritma SVM untuk menganalisis sentimen opini publik terhadap film "Agak Laen: Menyala Pantiku!" di platform X masih relatif terbatas.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengklasifikasikan opini

pengguna platform X mengenai film "Agak Laen: Menyala Pantiku!" ke dalam kategori sentimen positif, negatif, dan netral menggunakan pendekatan SVM. Selain itu, penelitian ini juga bertujuan untuk menganalisis kecenderungan sentimen publik terhadap film tersebut guna memperoleh gambaran yang lebih komprehensif mengenai persepsi masyarakat terhadap film "Agak Laen: Menyala Pantiku!" di media sosial.

Untuk mencapai tujuan tersebut, penelitian ini menerapkan pendekatan analisis sentimen dengan memanfaatkan teknik pemrosesan teks dan metode pembelajaran mesin. Data opini pengguna diperoleh dari platform X melalui proses *crawling*, kemudian diproses melalui tahapan *preprocessing* teks dan pembobotan fitur menggunakan metode TF-IDF. Selanjutnya, algoritma SVM digunakan untuk mengklasifikasikan sentimen opini pengguna ke dalam tiga kategori sentimen. Melalui pendekatan tersebut, penelitian ini diharapkan dapat memberikan gambaran mengenai distribusi dan kecenderungan sentimen publik terhadap film "Agak Laen: Menyala Pantiku!" di media sosial.

2. TINJAUAN PUSTAKA

2.1. *Text Mining*

Text mining merupakan proses mengolah data teks yang tidak terstruktur menjadi bentuk yang lebih terstruktur sehingga dapat dianalisis [8]. Teknik ini menggabungkan metode dari bidang *data mining, natural language processing* (NLP), dan *machine learning* untuk menemukan pola, hubungan, maupun informasi tersembunyi dalam dokumen teks [9]. Dalam perkembangannya, *text mining* telah banyak digunakan dalam berbagai bidang seperti analisis media sosial, analisis opini, sistem rekomendasi, hingga penelitian ilmiah berbasis bibliometrik.

Beberapa penelitian sebelumnya telah memanfaatkan teknik text mining untuk menganalisis data teks dalam berbagai bidang penelitian. Yang, Kinshuk, dan An menggunakan teknik *text mining* untuk menganalisis tren penelitian dalam bidang pendidikan dengan memanfaatkan metode klasifikasi dan analisis topik untuk mengidentifikasi pola penelitian dari sejumlah besar dokumen ilmiah [9]. Penelitian lain oleh Sandu dan rekan-rekannya menerapkan *text mining* untuk menganalisis data media sosial guna mengidentifikasi pola opini publik dan topik yang sering dibahas dalam platform digital [10]. Selain itu, Guindani dan rekan-rekannya menggunakan pendekatan *text mining* untuk menganalisis tren penelitian dalam bidang pertanian melalui analisis bibliometrik terhadap publikasi ilmiah [11]. Hasil dari berbagai penelitian tersebut menunjukkan bahwa *text mining* efektif digunakan untuk mengekstraksi informasi penting dari data teks yang berukuran besar dan tidak terstruktur.

Secara umum, proses *text mining* terdiri dari beberapa tahapan utama yang saling berkaitan, dimulai dari pengumpulan data teks (*data collection*) yang dapat berasal dari berbagai sumber seperti dokumen digital, artikel ilmiah, maupun media sosial. Selanjutnya dilakukan tahap *preprocessing* teks untuk membersihkan dan menormalisasi data melalui proses seperti *tokenization*, *stopword removal*, dan *stemming* sehingga kualitas data menjadi lebih baik untuk dianalisis [12]. Setelah itu, teks diubah ke dalam bentuk numerik melalui proses *feature extraction* atau representasi teks, misalnya menggunakan metode *Bag of Words* (BoW) atau *Term Frequency–Inverse Document Frequency* (TF-IDF), sehingga algoritma *machine learning* dapat memproses data tersebut. Tahap akhir melibatkan penerapan berbagai teknik analisis seperti *text classification*, *clustering*, *topic modeling*, atau *sentiment analysis* untuk menemukan pola maupun informasi yang relevan dari kumpulan data teks [13].

2.2. Analisis Sentimen

Analisis sentimen merupakan metode yang bertujuan menganalisis opini atau ulasan individu terhadap suatu topik tertentu. Teknik ini termasuk dalam bidang *text mining* dan *natural language processing* (NLP) yang bertujuan mengekstraksi informasi subjektif dari data tekstual seperti komentar, ulasan, atau postingan media sosial [14]. Dalam prosesnya, analisis sentimen dilakukan dengan mengklasifikasikan teks pada tingkat dokumen, kalimat, atau fitur tertentu ke dalam kategori polaritas sentimen seperti positif, negatif, atau netral [4]. Melalui proses tersebut, analisis sentimen dapat digunakan untuk menggambarkan kecenderungan opini publik secara sistematis serta membantu memahami persepsi masyarakat terhadap suatu produk, layanan, maupun isu tertentu [15], [16].

Metode analisis sentimen umumnya menggunakan berbagai pendekatan, seperti pendekatan berbasis leksikon (*lexicon-based*) dan pendekatan berbasis *machine learning*. Pendekatan leksikon memanfaatkan kamus kata yang telah diberi nilai sentimen untuk menentukan polaritas suatu teks, sedangkan pendekatan *machine learning* menggunakan algoritma klasifikasi seperti Naïve Bayes, *Support Vector Machine* (SVM), dan *Random Forest* untuk mempelajari pola sentimen dari data pelatihan [14], [17].

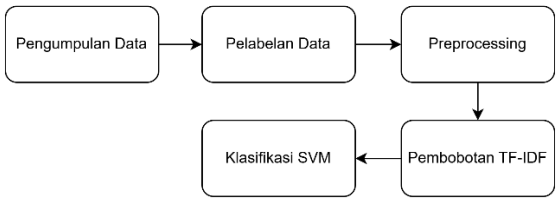
2.3. *Support Vector Machine*

Support Vector Machine (SVM) merupakan algoritma *machine learning* yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan data ke dalam kelas-kelas tertentu dengan margin maksimum [18]. SVM juga dikenal efektif dalam menangani data berdimensi tinggi, termasuk pada data dengan jumlah sampel kecil hingga menengah [19]. Kemampuan tersebut menjadikan SVM banyak digunakan dalam tugas klasifikasi teks, termasuk analisis sentimen. Hal ini didukung oleh penelitian [19] yang membuktikan bahwa penggabungan algoritma SVM dengan skema pembobotan TF-IDF sangat efektif dalam mengoptimalkan performa

klasifikasi sentimen positif dan negatif pada ulasan film Indonesia.

3. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan sistematis untuk membangun model analisis sentimen yang akurat. Tahapan penelitian dimulai dari pengumpulan data, pelabelan data, *preprocessing* data, *feature extraction*, hingga pembangunan model *machine learning*. Setiap tahapan dirancang secara terstruktur untuk memastikan kualitas data dan performa model yang optimal.



Gambar 1. Tahapan penelitian

3.1. Pengumpulan Data

Pengumpulan data merupakan tahapan awal dalam penelitian analisis sentimen ini. Data dikumpulkan melalui teknik *crawling*, yaitu proses pengambilan data secara otomatis dari platform digital menggunakan bahasa pemrograman Python dan *library* pendukung untuk mengekstraksi data teks dari sumber yang diteliti.

Penelitian ini melakukan proses *crawling* data dengan mengumpulkan *tweet* mengenai film Agak Laen 2 dari platform media sosial X (Twitter). Pengambilan data dilakukan dalam rentang waktu tanggal 10 November 2025 hingga tanggal 10 Februari 2026. Selama proses tersebut, sebanyak 2000 *tweet* berhasil dikumpulkan dan disimpan dalam format CSV. Data yang dikumpulkan mencakup informasi seperti teks *tweet*, tanggal posting, jumlah *likes*, dan jumlah *retweet*. Proses *crawling* dilakukan secara otomatis dengan *script* Python yang telah dirancang untuk mengekstraksi *tweet* yang mengandung kata kunci "Agak Laen 2", "Agak Laen: Menyala Pantiku".

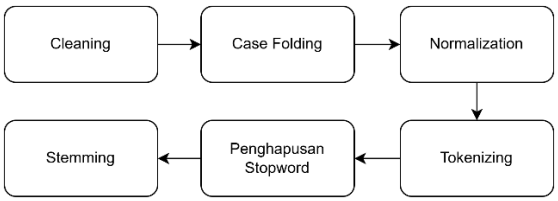
3.2. Pelabelan Data

Pelabelan Data merupakan tahap pemberian label sentimen pada setiap data teks yang digunakan dalam penelitian. Proses pelabelan dilakukan secara manual dengan membaca dan menilai setiap teks opini yang diperoleh dari platform X. Setiap data kemudian diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral.

Sentimen positif diberikan pada teks yang mengandung pujian, dukungan, atau penilaian baik terhadap film "Agak Laen: Menyala Pantiku!". Sentimen negatif diberikan pada teks yang berisi kritik, keluhan, atau penilaian buruk terhadap film tersebut. Sementara itu, sentimen netral diberikan pada teks yang bersifat informatif, tidak menunjukkan kecenderungan emosi tertentu, atau hanya menyampaikan fakta tanpa opini yang jelas. Hasil pelabelan ini digunakan sebagai data acuan dalam proses pelatihan dan pengujian model klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

3.3. *Preprocessing*

Preprocessing merupakan tahap penting dalam *pipeline text mining* untuk meningkatkan kualitas representasi fitur sebelum proses klasifikasi dilakukan.



Gambar 2. Tahapan preprocessing

a. *Cleaning*: proses penghapusan elemen yang tidak dibutuhkan dari teks mentah seperti URL, *username*, angka, tanda baca, emoji, dan simbol lain
b. *Case Folding*: proses untuk mengubah semua huruf menjadi huruf kecil (*lowercase*) sehingga variasi huruf besar dan kecil tidak dianggap berbeda oleh sistem.
c. *Normalization*: proses untuk menstandarkan bentuk kata, mengubah kata non-baku, singkatan, atau bahasa gaul menjadi bentuk baku yang sesuai dengan pedoman kamus bahasa (misalnya KBBI).
d. *Tokenizing*: proses untuk memecah kalimat atau dokumen teks menjadi unit-unit kecil, biasanya berupa kata atau token.
e. Penghapusan *Stopword*: tahap menghilangkan kata-kata umum yang sering muncul tetapi dianggap tidak memiliki makna signifikan dalam konteks klasifikasi teks, seperti "yang", "dan", "di", dsb.
f. *Stemming*: proses untuk mengubah kata menjadi bentuk dasar atau *root word* dengan menghapus imbuhan seperti awalan dan akhiran.

3.4. *Feature Extraction*

Feature extraction merupakan tahap dalam proses *text mining* yang bertujuan mengubah data teks menjadi representasi numerik sehingga dapat diproses oleh algoritma *machine learning* atau klasifikasi [20]. Tahap ini penting karena algoritma komputasi tidak dapat secara langsung memproses data dalam bentuk teks mentah [21], [22]. Pada penelitian ini, proses *feature extraction* dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) yang merupakan salah satu teknik pembobotan kata yang paling banyak digunakan dalam pengolahan teks dan information retrieval [23]. Metode ini terdiri dari dua komponen utama yaitu *Term Frequency* (TF) yang menghitung frekuensi kemunculan suatu kata

dalam sebuah dokumen dan *Inverse Document Frequency* (IDF) yang mengukur seberapa jarang kata tersebut muncul dalam keseluruhan kumpulan dokumen [24]. Kata yang sering muncul dalam suatu dokumen tetapi jarang muncul pada dokumen lain akan memperoleh bobot yang lebih tinggi sehingga dianggap lebih representatif dalam menggambarkan isi dokumen.

3.5. *Model Machine Learning*

Menurut [25], algoritma *Support Vector Machine* (SVM) mampu digunakan dalam analisis sentimen untuk mengklasifikasikan tanggapan publik dengan tingkat akurasi yang cukup stabil, meskipun masih memiliki keterbatasan dalam menangani variasi bahasa informal pada media sosial. Berdasarkan temuan tersebut, penelitian ini menggunakan algoritma SVM sebagai model machine learning utama untuk mengklasifikasikan sentimen data yang dikumpulkan, karena dinilai relevan dan sesuai dengan karakteristik data penelitian.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pada tahap ini dilakukan proses crawling data ulasan menggunakan bahasa pemrograman Python dengan kata kunci “Agak Laen 2” dan “Agak Laen: Menyala Pantiku!”. Proses pengambilan data dilakukan dalam rentang waktu yang telah ditentukan dan berhasil mengumpulkan sebanyak 2.000 *tweet* yang tersimpan dalam format CSV.

Tabel 1. Hasil crawling data

Text
nonton agak laen: menyala pantiku! sumpah deh gua ngakak sampe perut dan kepala gua pusing gegara kebanyakan ketawa bestttt weekend everrr djakarta theater penuh bgt besok nonton lagi di kota lain keren pakk @MuhadklyAcho https://t.co/OyhdtsMfuF
Agak laen menyala pantiku lebih lucu lagi jokenya lebih padet dan eksekusinya pas bgt menyala pantiku https://t.co/J7PiBhUWSN
@mmaryasir soalnya bikin bagan yg tokoh tokoh di agak laen menyala pantiku agak susah kalau blm tau ceritanya sih wkwk nice work pak! ditunggu karya di project lainnya xixi
Agak Laen menyala pantiku emang komedinya lebih darderdor daripada yang agak Laen 2024. Sepanjang nonton tenggorokan kering cape ketawa
Nonton #AgakLaen banyak banget bataknya euy. Mana kek kesentil soalnya boru Rajagukguk wkwk bagus asli bagus banget filmnya. Agak Laen 9/10 kalau Agak Laen: Menyala Pantiku 9.5/10. Semoga tembus 10juta penonton.

4.2. Pelabelan Data

Pelabelan data dilakukan secara manual untuk mengelompokkan *tweet* ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Hasil pelabelan menunjukkan bahwa mayoritas data termasuk dalam kategori positif sebanyak 1011 *tweet*, diikuti sentimen netral sebanyak 936 *tweet*, dan sentimen negatif sebanyak 53 *tweet.*

4.3. *Preprocessing* Data

Data mentah yang terkumpul masih mengandung berbagai komponen *noise* seperti URL, mention, emoji, *hashtag*, serta variasi bahasa tidak baku. Oleh karena itu, diperlukan tahap *preprocessing* sebelum dilakukan proses analisis lebih lanjut.

4.4. *Cleaning*

Proses ini difokuskan kepada penghapusan elemen yang tidak relevan bagi model, seperti URL, *mention*, angka, tanda baca, emoji, dan karakter non-alfabet.

Tabel 2. Hasil *cleaning*

Sebelum	Sesudah
Agak laen menyala pantiku lebih lucu lagi jokenya lebih padet dan eksekusinya pas bgt menyala pantiku https://t.co/J7PiBhUWSN	Agak laen menyala pantiku lebih lucu lagi jokenya lebih padet dan eksekusinya pas bgt menyala pantiku

4.5. *Case Folding*

Prosedur *case folding* diterapkan untuk menyeragamkan seluruh teks menjadi huruf kecil (*lowercase*). Langkah ini penting untuk mencegah terjadinya redundansi, di mana kata yang sama dianggap berbeda hanya karena perbedaan penggunaan huruf kapital.

Tabel 3. Hasil *case folding*

Sebelum	Sesudah
Agak laen menyala pantiku lebih lucu lagi jokenya lebih padet dan eksekusinya pas bgt menyala pantiku	agak laen menyala pantiku lebih lucu lagi jokenya lebih padet dan eksekusinya pas bgt menyala pantiku

4.6. *Normalization*

Kata tidak baku dan singkatan seperti “bgt” dan “padet” diubah menjadi bentuk baku “banget” dan “padat”. Tahap ini menghasilkan teks yang lebih konsisten dan sesuai dengan standar bahasa.

Tabel 4. Hasil *normalization*

Sebelum	Sesudah
agak laen menyala pantiku lebih lucu lagi jokenya lebih padet dan eksekusinya pas bgt menyala pantiku	agak laen menyala pantiku lebih lucu lagi jokenya lebih padat dan eksekusinya pas banget menyala pantiku

4.7. *Tokenizing*

Proses *tokenizing* dijalankan untuk memecah teks menjadi token atau kata-kata tunggal.

Tabel 5. Hasil *tokenizing*

Sebelum	Sesudah
agak laen menyala pantiku lebih lucu lagi jokenya lebih padet dan eksekusinya pas bgt menyala pantiku	['agak', 'laen', 'menyala', 'pantiku', 'lebih', 'lucu', 'lagi', 'jokenya', 'lebih', 'padat', 'dan', 'eksekusinya', 'pas', 'banget', 'menyala', 'pantiku']

4.8. Penghapusan *Stopword*

Kata umum yang tidak memiliki kontribusi terhadap makna sentimen seperti “dan”, “yang”, dan “di” dihapus.

Tabel 6. Hasil penghapusan *stopword*

Sebelum	Sesudah
['agak', 'laen', 'menyala', 'pantiku', 'lebih', 'lucu', 'lagi', 'jokenya', 'lebih', 'padat', 'dan', 'eksekusinya', 'pas', 'banget', 'menyala', 'pantiku']	agak laen menyala pantiku lucu jokenya padat eksekusinya pas banget menyala pantiku

4.9. *Stemming*

Proses *stemming* dilakukan dengan mengonversi setiap kata berimbuhan kembali ke akar katanya. Hasil *stemming* mengurangi variasi bentuk kata sehingga dimensi fitur menjadi lebih sederhana.

Tabel 7. Hasil *stemming*

Sebelum	Sesudah
agak laen menyala pantiku lucu jokenya padat eksekusinya pas banget menyala pantiku	agak laen nyala panti lucu jokenya padat eksekusi pas banget nyala panti

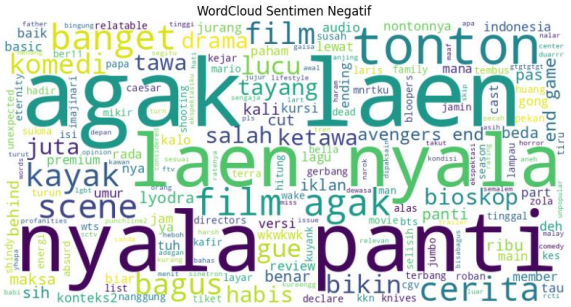
4.10. *Word Cloud*

Representasi visual melalui *word cloud* diimplementasikan untuk mengidentifikasi frekuensi kemunculan kata yang paling signifikan pada tiap-tiap label sentimen.

Gambar 3. *Wordcloud* sentimen positif

Analisis menunjukkan bahwa pada sentimen positif, kata-kata seperti “lucu”, “bagus”, “keren”, dan “ketawa” muncul dengan frekuensi tinggi. Hal ini mengindikasikan bahwa aspek komedi menjadi daya tarik utama film di mata penonton.

Sementara itu, pada sentimen negatif, kata-kata yang muncul cenderung berkaitan dengan kritik terhadap alur cerita atau ekspektasi yang tidak terpenuhi. Walaupun jumlahnya relatif kecil, keberadaan sentimen negatif tetap memberikan gambaran bahwa tidak semua penonton memiliki pengalaman yang sama.



Gambar 4. *Wordcloud* sentimen negatif

4.11. TF-IDF

Berdasarkan hasil proses *feature extraction* menggunakan metode TF-IDF, diperoleh sejumlah kata yang memiliki frekuensi kemunculan tertinggi dalam dataset opini pengguna.

	Word	Frequency
0	agak	2244
1	laen	2221
2	nyala	2060
3	panti	2022
4	tonton	993
5	film	829
6	banget	457
7	lucu	250
8	juta	241
9	ketawa	202

Gambar 5. Kata dengan frekuensi tertinggi

Kemunculan kata-kata seperti lucu dan ketawa mengindikasikan bahwa banyak opini pengguna menyoroti unsur komedi yang menjadi karakteristik utama film tersebut. Secara keseluruhan, hasil pembobotan kata melalui TF-IDF membantu mengidentifikasi kata-kata penting yang merepresentasikan isi teks dan selanjutnya digunakan sebagai fitur dalam proses klasifikasi sentimen menggunakan algoritma *Support Vector Machine* (SVM).

4.12. Support Vector Machine

Classification Report:

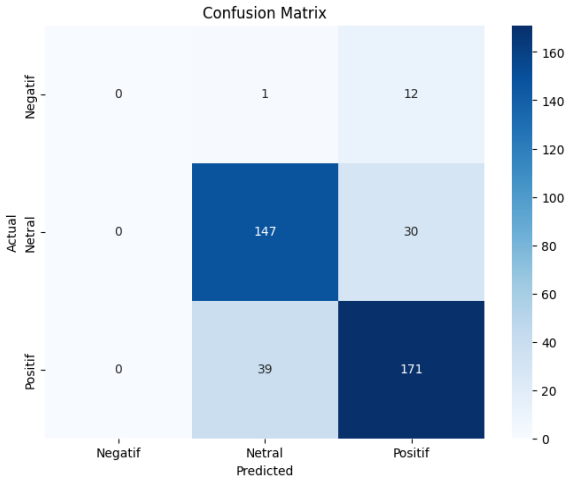
	precision	recall	f1-score	support
Negatif	0.00	0.00	0.00	13
Netral	0.79	0.83	0.81	177
Positif	0.80	0.81	0.81	210
accuracy			0.80	400
macro avg	0.53	0.55	0.54	400
weighted avg	0.77	0.80	0.78	400

Gambar 6. *Classification report*

Dataset yang telah melalui tahap preprocessing, kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan metode train-test split dari

Scikit-learn, serta diubah ke dalam representasi numerik menggunakan TF-IDF melalui TfidfVectorizer. Kinerja model kemudian dievaluasi menggunakan metrik *accuracy*, *confusion matrix*, dan *classification report* yang mencakup *precision*, *recall*, dan *F1-score*.

Berdasarkan hasil pengujian, model SVM memperoleh tingkat akurasi sebesar 0,80 (80%) dari total 400 data uji. Hasil evaluasi tersebut mengonfirmasi efektivitas SVM dalam membedakan kategori sentimen netral dan positif secara presisi. Di sisi lain, performa model dalam mengidentifikasi sentimen negatif ditemukan belum optimal. Kondisi tersebut dipengaruhi oleh ketidakseimbangan jumlah data pada setiap kelas, mengingat opini dengan sentimen positif dan netral jauh lebih dominan dibandingkan sentimen negatif dalam dataset.



Gambar 7. *Confusion matrix*

Berdasarkan *confusion matrix*, dari total 400 data uji terlihat bahwa model memiliki kecenderungan dalam mengklasifikasikan data ke dalam kelas netral dan positif. Pada kelas negatif yang berjumlah 13 data, tidak terdapat data yang berhasil diprediksi sebagai sentimen negatif. Sebanyak 12 data diklasifikasikan sebagai positif, sedangkan 1 data lainnya diprediksi sebagai netral. Kemampuan model dalam mengenali sentimen negatif masih sangat rendah.

Dari total 177 data pada kelas netral, sebanyak 147 data berhasil diprediksi dengan benar sebagai sentimen netral, sedangkan 30 data lainnya salah diklasifikasikan sebagai sentimen positif. Hasil ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengenali karakteristik sentimen netral. Namun, masih terdapat kecenderungan kesalahan klasifikasi ke dalam kelas positif yang mengindikasikan adanya kemiripan pola fitur antara kedua kelas tersebut.

Pada kategori netral, model secara akurat memprediksi 147 dari total 177 data ulasan. Meskipun demikian, tercatat sebanyak 30 data mengalami misklasifikasi ke dalam label positif. Hasil ini menunjukkan bahwa model cukup mampu mengenali sentimen positif, namun masih terdapat kesalahan klasifikasi terhadap kelas netral.

Secara keseluruhan, confusion matrix menunjukkan bahwa kesalahan klasifikasi terbesar terjadi pada pergeseran antara kelas netral dan positif, serta kegagalan model dalam mengenali kelas negatif. Hal ini disebabkan adanya ketidakseimbangan distribusi data pada dataset, di mana jumlah data negatif jauh lebih sedikit dibandingkan kelas lainnya, sehingga memengaruhi kemampuan model dalam mempelajari karakteristik sentimen negatif.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, penerapan model *Support Vector Machine* (SVM) dengan fitur TF-IDF mampu melakukan klasifikasi sentimen terhadap opini penonton film dengan cukup baik. Model yang dibangun menghasilkan akurasi sebesar 80% pada data uji, dengan performa yang relatif baik dalam mengenali sentimen netral dan positif. Namun, model masih kurang optimal dalam mendeteksi sentimen negatif karena jumlah data pada kelas tersebut jauh lebih sedikit dibandingkan kelas lainnya, sehingga menyebabkan ketidakseimbangan data yang memengaruhi proses pembelajaran model. Selain itu, hasil analisis kata menunjukkan bahwa opini penonton didominasi oleh kata-kata seperti lucu, ketawa, dan bagus, yang mengindikasikan bahwa unsur komedi menjadi faktor utama yang banyak disoroti oleh penonton dalam memberikan tanggapan terhadap film tersebut.

Berdasarkan kondisi tersebut, ketidakseimbangan jumlah data antar kelas sentimen perlu menjadi perhatian pada penelitian selanjutnya. Penerapan teknik penyeimbangan data dapat dipertimbangkan untuk mengatasi perbedaan distribusi data antar kelas. Salah satu metode yang dapat digunakan adalah SMOTE (*Synthetic Minority Over-sampling Technique*) yang berfungsi menghasilkan data sintetis pada kelas minoritas sehingga jumlahnya meningkat. Dengan penerapan metode tersebut, diharapkan distribusi data antar kelas menjadi lebih seimbang sehingga model klasifikasi dapat mempelajari karakteristik setiap kelas dengan lebih baik dan meningkatkan kemampuan dalam mendeteksi sentimen negatif.

DAFTAR PUSTAKA

[1] A. R. Bali, “DAHSYAT! Film Agak Laen 2 Sudah Tembus Penonton 11 Juta Lebih, Sukses Tendang Rekor Film Hollywood, Avengers Endgame yang Cuma 10.976.338!,” radarbali.id. Accessed: Mar. 01, 2026. [Online]. Available: https://radarbali.jawapos.com/hiburan-budaya/707192717/dahsyat-film-agak-laen-2-sudah-tembus-penonton-11-juta-lebih-sukses-tendang-rekor-film-hollywood-avengers-endgame-yang-cuma-10976338

[2] R. R. Ilmi, F. Kurniawan, and S. Harini, “Prediksi Rating Film Imdb Menggunakan Decision Tree,”

J. Teknol. Inf. dan Ilmu Komput., vol. 10, no. 4, pp. 791–798, 2023, doi: https://doi.org/10.25126/jtiik.2024106615.

[3] A. L. Afriani, D. Suprayitno, and N. A. Misbah, "Pengaruh Media Online Reviews Terhadap Keputusan Menonton Film," *J. Penelitan Sos. Ilmu Komun.*, vol. 7, no. 1, pp. 1–10, 2023, doi: https://doi.org/10.33751/jpsik.v7i1.7573.

[4] 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," *J. Teknol. dan Komun. Pemerintah.*, vol. 4, no. 1, pp. 1–13, 2022, doi: https://doi.org/10.33701/jtkp.v4i1.2701.

[5] S. B. Putri, Y. N. Anisa, and N. Saputra, "Analisis Sentimen Film Kuliah Kerja Nyata (Kkn) Di Desa Penari Menggunakan Metode Naive Bayes," *J. Sist. Teknol. Inf. Komun.*, vol. 5, no. 2, pp. 22–26, 2022, doi: https://doi.org/10.32524/jusitik.v5i2.704.

[6] A. R. Se, R. Amarullah, and M. R. Pribadi, "Analisis Sentimen Masyarakat Terhadap Film Animasi Jumbo Di Platform Tiktok Menggunakan Algoritma Naïve Bayes," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 3, pp. 181–187, 2025, doi: https://doi.org/10.23960/jitet.v13i3.6669.

[7] M. D. Aulia and Y. Akbar, "Penerapan Algoritma Naive Bayes dalam Sistem Analisis Sentimen Media Sosial X terhadap Film Agak Laen," *J. Indones. Manaj. Inform. dan Komun.*, vol. 6, no. 3, pp. 1958–1967, 2025, doi: https://doi.org/10.63447/jimik.v6i3.1608.

[8] S. A. S. Mola, R. V. K. I. O. Roma, and T. Widiastuti, *Text Mining Analisis Sentimen Dengan Lexicon*, 1st ed. Bandung: Kaizen Media Publishing, 2025.

[9] J. Yang, Kinshuk, and Y. An, "A survey of the literature: how scholars use text mining in Educational Studies?," *Educ. Inf. Technol.*, vol. 28, no. 2, pp. 2071–2090, 2023, doi: 10.1007/s10639-022-11193-3.

[10] A. Sandu, L. Cotfas, A. Stanescu, and C. Delcea, "A Bibliometric Analysis of Text Mining : Exploring the Use of Natural Language Processing in Social Media Research," *Appl. Sci.*, vol. 14, no. 8, pp. 1–34, 2024, doi: https://doi.org/10.3390/app14083144.

[11] L. G. Guindani, G. A. Oliveirai, M. H. Dal, M. Ribeiro, G. V. Gonzalez, and J. D. de Lima, "Exploring current trends in agricultural commodities forecasting methods through text mining : Developments in statistical and artificial intelligence methods," *Heliyon*, vol. 10, no. 23, pp. 1–16, 2024, doi: 10.1016/j.heliyon.2024.e40568.

[12] Q. Xu, H. Gu, and S. Ji, "Text clustering based on pre-trained models and autoencoders," *Front. Comput. Neurosci.*, vol. 17, pp. 1–13, 2022, doi: https://doi.org/10.3389/fncom.2023.1334436.

[13] S. Xiong, W. Tian, H. Si, G. Zhang, and L. Shi, "A Survey of the Applications of Text Mining for the Food Domain," *Algorithms*, vol. 17, no. 5, pp. 1–223, 2024, doi: https://doi.org/10.3390/a17050176.

[14] M. Kumar, L. Khan, and H. Chang, "Evolving techniques in sentiment analysis : a comprehensive review," *Peer Comput. Sci.*, vol. 11, pp. 1–61, 2025, doi: https://doi.org/10.7717/peerj-cs.2592.

[15] T. Anderson, S. Sarkar, and R. Kelley, "Analyzing public sentiment on sustainability : A comprehensive review and application of sentiment analysis techniques," *Nat. Lang. Process. J.*, vol. 8, no. 1, p. 100097, 2024, doi: https://doi.org/10.1016/j.nlp.2024.100097.

[16] J. Min, E. K. Alvin, Y. Lei, and L. Y. Chong, "Reinforcement learning in sentiment analysis : a review and future directions," *Artif. Intell. Rev.*, vol. 58, no. 6, pp. 1–40, 2025, doi: https://doi.org/10.1007/s10462-024-10967-0 Reinforcement.

[17] T. A. Al-Qablan, M. H. Mohd Noor, M. A. Al-Betar, and A. T. Khader, "A survey on sentiment analysis and its applications," *Neural Comput. Appl.*, vol. 35, no. 29, pp. 21567–21601, 2023, doi: 10.1007/s00521-023-08941-y.

[18] D. R. Nurqotimah, A. N. Khudori, and R. S. Pradini, "Implementasi Algoritma Support Vector Machine (SVM) Untuk Klasifikasi Penyakit Stroke," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 2, pp. 179–185, 2024, doi: https://doi.org/10.52158/jacost.v5i2.817.

[19] J. B. S. Emarapenta and H. C. K. Sitorus, "Analisis Sentimen Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF Sentiment Analysis of Public towards Indonesian Horror Films Using SVM and TF- IDF Methods," *J. Manaj. Inform.*, vol. 14, no. 1, pp. 42–53, 2024, doi: https://doi.org/10.34010/jamika.v14i1.11946.

[20] F. Indrajid and I. N. S. W. Wijaya, "MODEL KLASIFIKASI TWEET TERKAIT ISU #INDONESIAGELAP MENGGUNAKAN SUPPORT VECTOR MACHINE BERBASIS DISTRIBUSI TOPIK LDA," *J. Inform. dan Tek. Elektro Terap.*, vol. 14, no. 1, pp. 934–944, 2026, doi: http://dx.doi.org/10.23960/jitet.v14i1.8878.

[21] M. Karim *et al.*, "Comprehension of polarity of articles by citation sentiment analysis using TF-IDF and ML classifiers," *Peer Comput. Sci.*, vol. 7, pp. 1–21, 2022, doi: 10.7717/peerj-cs.1107.

[22] M. M. Danyal, S. S. Khan, M. Khan, S. Ullah, M. B. Ghaffar, and W. Khan, "Sentiment analysis of movie reviews based on NB approaches using TF–IDF and count vectorizer," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, p. 87, 2024, doi: 10.1007/s13278-024-01250-9.

[23] L. Xiao, Q. Li, Q. Ma, J. Shen, Y. Yang, and D. Li, "Text classification algorithm of tourist attractions subcategories with modified TF- IDF and Word2Vec," *PLoS One*, vol. 19, no. 10, pp. 1–34, 2024, doi: https://doi.org/10.1371/journal.pone.0305095.

[24] O. I. Gifari, M. Adha, I. R. Hendrawan, F. Freddy, and S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," vol. 2, no. 1, pp. 36–40, 2022.

[25] M. H. Arfian *et al.*, "Analisis Sentimen Pada Media Sosial Menggunakan Metode Support Vector Machine," *J. Ilmu Tek. dan Komput.*, vol. 09, no. 01, pp. 1–6, 2025, doi: 10.22441/jitkom.v9i1.001.