

KOMPARASI ANALISIS SENTIMEN ULASAN FILM AVENGERS:
ENDGAME DI IMDB MENGGUNAKAN METODE NAÏVE BAYES DAN
SVM

¹⁾Radendha Muhammad Arthansa, ²⁾Dhea Intan Sagita, ^{3*)}Anggraini Puspita Sari
^{1,2,3)}Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur
¹⁾22081010218@student.upnjatim.ac.id, ²⁾22081010016@student.upnjatim.ac.id,
^{3*)}anggraini.puspita.if@upnjatim.ac.id

INFO ARTIKEL	ABSTRAK
Riwayat Artikel : Diterima : 14 Agustus 2024 Disetujui : 18 Agustus 2024	Kualitas sebuah film seringkali dievaluasi melalui opini atau ulasan dari penonton yang dapat diklasifikasikan menjadi sentimen positif dan negatif. Penelitian ini bertujuan untuk menginvestigasi respons penonton terhadap film “Avengers: Endgame” di IMDb menggunakan dua algoritma <i>data mining</i> yang populer, yaitu Naïve Bayes dan Support Vector Machine (SVM). Topik ini dipilih karena kepopuleran dan dampak global film tersebut, serta kebutuhan untuk memahami pandangan penonton secara mendalam. Dengan menggunakan 1500 ulasan sebagai dataset, hasil analisis menunjukkan bahwa algoritma SVM memiliki akurasi sebesar 83%, sedikit lebih unggul dibandingkan dengan Naïve Bayes yang memiliki akurasi sebesar 82%. Temuan ini memberikan gambaran yang lebih komprehensif tentang bagaimana penonton merespons film “Avengers: Endgame” IMDb. Meskipun SVM menunjukkan akurasi yang lebih tinggi, akurasi 82% dari Naïve Bayes juga menunjukkan keandalan dan validitas metode ini dalam analisis sentiment ulasan film.
Kata Kunci : Analisis Sentimen, Komparasi, Avengers: Endgame, Naïve Bayes, SVM	

ARTICLE INFO	ABSTRACT
Article History : Received : August 14, 2024 Accepted : August 18, 2024	<i>The quality of a movie is often evaluated through audience opinions or reviews, which can be classified into positive and negative sentiments. This study aims to investigate audience responses to the movie "Avengers: Endgame" on IMDb using two popular data mining algorithms, namely Naïve Bayes and Support Vector Machine (SVM). This topic was chosen due to the global popularity and impact of the movie, as well as the need to understand audience views in depth. Using 1500 reviews as the dataset, the analysis results show that the SVM algorithm has an accuracy of 83%, slightly superior to Naïve Bayes which has an accuracy of 82%. These findings provide a more comprehensive picture of how audiences responded to the movie "Avengers: Endgame" movie on IMDb. Although SVM shows higher accuracy, the 82% accuracy of Naïve Bayes also shows the reliability and validity of this method in sentiment analysis of movie reviews.</i>
Keywords: Sentiment Analysis, Comparison, Avengers: Endgame, Naïve Bayes, SVM	

1. PENDAHULUAN

Pada era digital, situs ulasan film telah menjadi media populer bagi masyarakat untuk menilai kualitas sebuah film. Emosi dalam kritik, komentar, atau umpan balik memberikan indikator yang memiliki tujuan berbeda dan dapat dikategorikan berdasarkan popularitas. Sumber hiburan yang paling disukai masyarakat umum adalah film. Ada banyak sekali bisnis produksi global di zaman globalisasi ini. Contohnya *Pixar*, *The Walt Disney Company*, *Warner Bros*, *Sony Pictures*, dan *Marvel Entertainment* (Frando Febriant dkk., 2023).

Marvel Entertainment memanfaatkan karakter-karakter waralabanya di seluruh dunia melalui berbagai sektor seperti hiburan, lisensi, dan penerbitan. *Marvel Studios*, sebuah perusahaan produksi film dan televisi Amerika, menciptakan *Marvel Cinematic Universe* (MCU), yang merupakan bagian dari *Marvel Entertainment* (Rifka Pratama, 2022). *Marvel Cinematic Universe* (MCU) telah menciptakan sejumlah film terkenal yang menarik minat publik, salah satunya adalah *Avengers: Endgame*.

Film “*Avengers: Endgame*”, yang dirilis pada tahun 2019, merupakan film ke-22 dalam *Marvel Cinematic Universe* (MCU) dan menjadi puncak dari narasi panjang yang dimulai dari film “*Iron Man*” pada tahun 2008. “*Avengers: Endgame*”, sebagai salah satu film yang paling dinanti dan memiliki basis penggemar yang sangat luas, menarik perhatian luar biasa dari para penonton dan kritikus di seluruh dunia. Film ini memecahkan rekor sebagai film dengan pendapatan tertinggi sepanjang masa.

Film tersebut menerima berbagai ulasan, kritikan, dan diskusi yang intens di berbagai platform online. *Internet Movie Database* atau yang lebih dikenal dengan IMDb, merupakan sebuah situs web yang berkaitan dengan film dan pembuatan film. Ulasan pengguna dan halaman peringkat adalah sumber utama informasi di situs web. Situs web ini menggunakan sistem peringkat Dimana skor 1 mewakili ulasan negative dan skor 10 mewakili ulasan positif (Gunawan dkk., 2022).

Situs web IMDb didedikasikan untuk film dan pembuatan film, sinopsis film, tanggal rilis di negara-negara tertentu, dan ulasan pengguna semuanya disertakan. Pendapat orang lain dan

ulasan film biasanya berdampak pada keputusan seseorang untuk membeli atau menonton film (Chandani dan Wahono, 2015).

Mengingat pentingnya ulasan ini, analisis sentimen menjadi alat yang berharga untuk memahami bagaimana film tersebut diterima oleh penonton. Analisis sentimen adalah metode untuk mengelompokkan komentar pengguna ke dalam kategori positif, negatif, atau netral (Rahman dkk., 2024). Analisis data sentimen merupakan prosedur yang rumit dan terdiri dari beberapa langkah yang memerlukan lima proses yang berbeda.

Dalam penelitian ini, dua metode utama yang digunakan untuk menganalisis sentimen film adalah Naïve Bayes dan Support Vector Machine (SVM). Hasil analisis sentimen yang optimal bisa didapat dengan menggunakan algoritma untuk perbandingan (Ardana dan Kambau, 2023).

Kedua metode tersebut dipilih karena keduanya telah digunakan secara luas dalam berbagai aplikasi NLP lainnya, seperti *text classification* dan *sentiment analysis*. Naïve Bayes digunakan karena algoritma sederhana tetapi nilai akurasi cukup tinggi. Berbeda dengan SVM yang dapat mengidentifikasi *hyperplane* yang berbeda yang meningkatkan kesenjangan antara dua kelas (Agustin, 2024).

Pada penelitian sebelumnya yang berjudul “*The Sentiment Analysis of Spider-Man: No Way Home Film Based on IMDb Reviews*” yang dilakukan pada tahun 2022, penelitian tersebut menunjukkan keefektifan berbagai teknik *machine learning* dalam analisis sentimen, seperti Naïve Bayes, Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), dan Decision Tree. Selain itu, penelitian ini memberikan wawasan tentang opini publik mengenai film “*Spider-Man: No Way Home*.” Hasilnya menunjukkan akurasi tertinggi dengan pengklasifikasi SVM dan SGD, yaitu akurasi sebesar 82% dan *f1-score* sebesar 81%.

Dengan latar belakang tersebut, penulis melakukan penelitian untuk membandingkan dua metode klasifikasi dalam analisis sentimen dengan objek penelitian yang berbeda. Tujuan penelitian ini adalah untuk memastikan pendekatan mana yang lebih bekerja lebih baik dalam menilai ulasan film. Untuk

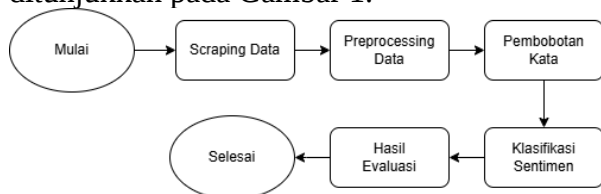
membandingkan hasil analisis, kedua metode diuji pada dataset yang sama.

Diharapkan bahwa hasil analisis ini dapat berkontribusi pada pengembangan sistem analisis sentimen yang menganalisis ulasan film dengan lebih akurat dan efektif serta memberikan bahan referensi yang lebih luas dan membantu dalam pengembangan sistem analisis sentimen yang lebih baik.

2. METODE

2.1. Alur Penelitian

Bagian ini akan memberikan penjelasan mengenai perancangan sistem dalam penelitian. Metode dalam penelitian ini diimplementasikan dalam beberapa langkah. Secara umum, alur penelitian dapat dijelaskan melalui diagram yang ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

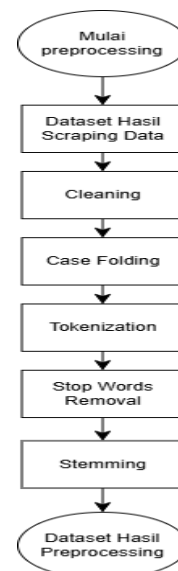
2.2. Scraping Data

Tahapan ini melibatkan proses pengambilan data ulasan film “*Avengers: Endgame*” dari situs web IMDb. Terdapat total 2003 ulasan yang digunakan sebagai data untuk penelitian ini yang dikumpulkan dari situs web IMDb dengan teknik scraping menggunakan aplikasi Octoparse 8.

Proses scraping dimulai dengan mengidentifikasi elemen-elemen HTML yang relevan pada halaman ulasan. Aplikasi Octoparse 8 kemudian digunakan untuk mengatur dan menjalankan *script scraping* secara otomatis memastikan bahwa semua ulasan yang relevan disimpan dalam format *comma-separated values* (CSV).

2.3. Preprocessing Data

Pada langkah ini, data teks yang tidak terstruktur diubah menjadi teks yang dapat diproses oleh komputer. Proses ini diperlukan karena kalimat atau teks yang diekstrak dari sumbernya merupakan data mentah, sehingga lebih sulit untuk diproses oleh komputer (Arif Widiawan Subagio dkk., 2024).



Gambar 2. Alur Preprocessing Data

Langkah preprocessing yang ditunjukkan pada Gambar 2, meliputi:

1. *Cleaning* adalah proses penghapusan karakter yang tidak diperlukan seperti angka, tanda baca, URL, dan tag HTML. Hal ini bertujuan untuk memastikan bahwa hanya teks ulasan yang bersih dan relevan yang digunakan dalam analisis (Arif Widiawan Subagio dkk., 2024).
2. *Case folding* adalah proses konversi huruf kapital menjadi huruf kecil (Darwis dkk., 2020). Langkah ini dilakukan untuk memastikan konsistensi dalam pengolahan teks, perlu diingat bahwa komputer dapat membedakan antara huruf besar dan huruf kecil.
3. *Tokenization* adalah proses memecah kalimat teks menjadi unit-unit kata yang dikenal sebagai token (Arif Widiawan Subagio dkk., 2024).
4. *Stop words removal* adalah proses menghapus kata-kata umum yang dianggap kurang signifikan dan muncul secara teratur dalam jumlah besar. Proses ini membantu mengurangi kebisingan data (Arif Widiawan Subagio dkk., 2024).
5. *Stemming* adalah proses menghilangkan awalan dan akhiran dari kata-kata yang mengandung kata sambung, kata depan, atau kata ganti untuk membuat kata dasar (Darmawan dkk., 2023).

2.4. Pembobotan Kata

Pembobotan kata merupakan langkah penting untuk mengubah teks ulasan menjadi representasi numerik. Metode ini menggunakan teknik *Count vectorizer* (CV). *Count vectorizer* berfungsi untuk menentukan nilai tingkat token. *Count vectorizer* mencatat frekuensi setiap token unik dalam dokumen. Semakin tinggi nilai token, semakin sering token tersebut muncul dalam teks (Sajid Khan dkk., 2023).

Count vectorizer memvektorisasi kalimat menggunakan library *scikit-learn*. Dengan menggunakan kata-kata dari setiap kalimat, modul tersebut akan menyusun daftar kosakata dari semua istilah berbeda yang digunakan dalam kalimat (Assegaff dkk., 2023).

2.5. Klasifikasi Sentimen

Tahap klasifikasi sentimen bertujuan untuk mengkategorikan ulasan film “*Avengers: Endgame*” dari IMDb ke dalam kelas sentimen positif atau negatif. Penelitian ini menggunakan klasifikasi Naïve Bayes dan SVM.

1. Klasifikasi Naïve Bayes

Klasifikasi Naïve Bayes ini menggunakan salah satu *class* dalam library *scikit-learn* di *Python* yaitu *MultinomialNB*. Karena klasifikasi Naïve Bayes semacam ini dapat menangani data teks, maka digunakanlah *MultinomialNB* (Noviana dan Rasal B A Jurusan, 2023).

2. Klasifikasi SVM

Terdapat dua jenis SVM yang berbeda, yaitu SVM non-linear dan SVM linear. SVM non-linear memetakan ke ruang dimensi yang lebih tinggi menggunakan fungsi trik kernel, sedangkan SVM linear membagi dua *class* pada *hyperplane* dengan batas-batas yang *soft* (Gracezando Yehova dan Ihsan Jambak, 2024).

2.6. Evaluasi Model

Tujuan dari tahap evaluasi model untuk menilai kinerja algoritma klasifikasi yang dilakukan dalam penelitian (Nurian, Padilah dan Garno, 2024). Evaluasi dilakukan dengan data uji, yaitu data yang berbeda dengan data latih yang digunakan untuk melatih model. Tahapan ini menggunakan *confusion matrix*, adalah

teknik penilaian yang digunakan untuk mengukur seberapa baik kinerja metode klasifikasi (Darmawan dkk., 2023). Hasil dari prosedur klasifikasi diwakili oleh empat nilai yang dihasilkan dalam tabel *confusion matrix*, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Ilustrasi tabel dapat dilihat pada Tabel.

Tabel 1. Tabel Perhitungan Confusion Matrix

		Nilai Prediksi	
		Positif	Negatif
Nilai Aktual	Positif	TP	FN
	Negatif	FP	TN

Hasil nilai prediksi model yang dapat digunakan untuk menghasilkan evaluasi metrik adalah *accuracy*, *precision*, *recall*, dan *f1-score*.

1. *Accuracy* untuk memastikan seberapa dekat nilai yang diharapkan dan nilai actual menyerupai satu sama lain (Nurian dkk., 2024).
2. *Precision* untuk menentukan ketepatan atau akurasi model berdasarkan hasil positif. Untuk menilai apakah nilai *False Positive* pada suatu model tinggi (Nurian dkk., 2024).
3. *Recall* untuk menentukan jumlah nilai *Actual Positive* yang dapat diklasifikasi oleh model sebagai *True Positive* dengan benar. Ketika terdapat nilai *False Negative* yang tinggi, *recall* merupakan statistik lain yang digunakan untuk memilih model yang optimal (Nurian dkk., 2024).
4. *F1-score* adalah rasio nilai rata-rata dari *recall* dan *precision* (Nurian dkk., 2024).

3. HASIL DAN PEMBAHASAN

3.1. Scraping Data

Data ulasan film *Avengers: Endgame* berhasil diambil dari situs web IMDb menggunakan teknik web scraping dengan bantuan perangkat lunak Octoparse 8. Penggunaan perangkat lunak ini memungkinkan pengumpulan sejumlah besar ulasan dengan efisien dan akurat. Data ulasan yang terkumpul kemudian disimpan untuk tahap preprocessing selanjutnya. Contoh hasil scraping tersedia dalam Tabel 2.

Tabel 2. Contoh Dokumen Hasil Scraping

Dokumen	Ulasan
Dokumen 1	“A 3 hours Marvel film with 2.5 hours of drama, 15 mins of actions and 15 mins of something else.. seriously?Their logic didn’t make any sense.Besides, what they did to Thor was brutal.”
Dokumen 2	“I LOVE MARVEL!!! This movie is incredible!! I just wish Stan lee was here to see everything he had worked to create!!”
Dokumen 3	“No Negative reviews.. This is one of my most favourite movies of 2019”

3.2. Preprocessing Data

Data yang telah dikumpulkan selama proses scraping disiapkan sebelum pengklasifikasian. Langkah preprocessing ini diperlukan sebelum analisis data karena prosedur ini mencoba mengubah data yang tidak terstruktur menjadi data yang terstruktur, yang akan membuat analisis menjadi lebih mudah dan mengurangi kesalahan (Gracezando Yehova dan Ihsan Jambak, 2024). Terdapat 5 tahapan yang dilakukan yaitu:

1. Cleaning

Tabel 3. Proses Cleaning

Ulasan Awal	Sesudah Cleaning
“A 3 hours Marvel film with 2.5 hours of drama, 15 mins of actions and 15 mins of something else.. seriously?Their logic didn’t make any sense.Besides, what they did to Thor was brutal.”	A hours Marvel film with hours of drama mins of actions and mins of something else seriouslyTheir logic didnt make any senseBesides what they did to Thor was brutal
“I LOVE MARVEL!!! This movie is incredible!! I just wish Stan lee was here to see everything he had worked to create!!”	I LOVE MARVEL This movie is incredible I just wish Stan lee was here to see

here to see everything he had worked to create!!”	everything he had worked to create
“No Negative reviews.. This is one of my most favourite movies of 2019”	No Negative reviews This is one of my most favourite movies of 2019”

2. Case Folding

Tabel 4. Proses Case Folding

Ulasan Awal	Sesudah Case Folding
“A 3 hours Marvel film with 2.5 hours of drama, 15 mins of actions and 15 mins of something else.. seriously?Their logic didn’t make any sense.Besides, what they did to Thor was brutal.”	a hours marvel film with hours of drama mins of actions and mins of something else seriously their logic didnt make any sense besides what they did to thor was brutal
“I LOVE MARVEL!!! This movie is incredible!! I just wish Stan lee was here to see everything he had worked to create!!”	i love marvel this movie is incredible i just wish stan lee was here to see everything he had worked to create
“No Negative reviews.. This is one of my most favourite movies of 2019”	no negative reviews this is one of my most favourite movies of 2019”

3. Tokenization

Tabel 5. Proses Tokenization

Ulasan Awal	Sesudah Tokenization
“A 3 hours Marvel film with 2.5 hours of drama, 15 mins of actions and 15 mins of something else.. seriously?Their logic didn’t make	['a', 'hours', 'film', 'marvel', 'film', 'with', 'hours', 'of', 'drama', 'mins', 'of', 'actions', 'and', 'mins', 'of', 'something', 'else', 'seriously', 'Their', 'logic', 'didn't', 'make

any sense.Besides, what they did to Thor was brutal.”	'seriouslytheir', 'logic', 'didnt', 'make', 'any', 'sensebesides', 'what', 'they', 'did', 'to', 'thor', 'was', 'brutal']
“I LOVE MARVEL!!! This movie is incredible!! I just wish Stan lee was here to see everything he had worked to create!!”	['i', 'love', 'marvel', 'this', 'movie', 'is', 'incredible', 'i', 'just', 'wish', 'stan', 'lee', 'was', 'here', 'to', 'see', 'everything', 'he', 'had', 'worked', 'to', 'create']
“No Negative reviews.. This is one of my most favourite movies of 2019”	['no', 'negative', 'reviews', 'this', 'is', 'one', 'of', 'my', 'most', 'favourite', 'movies', 'of']

4. Stop Words Removal

Tabel 6. Proses Stop Words Removal

Ulasan Awal	Sesudah Stop Words Removal
“A 3 hours Marvel film with 2.5 hours of drama, 15 mins of actions and 15 mins of something else.. seriously?Their logic didn't make any sense.Besides, what they did to Thor was brutal.”	['hours', 'marvel', 'film', 'hours', 'drama', 'mins', 'actions', 'mins', 'something', 'else', 'seriouslytheir', 'logic', 'didnt', 'make', 'sensebesides', 'thor', 'brutal']
“I LOVE MARVEL!!! This movie is incredible!! I just wish Stan lee was here to see everything he had worked to create!!”	['love', 'marvel', 'movie', 'incredible', 'wish', 'stan', 'lee', 'see', 'everything', 'worked', 'create']
“No Negative reviews.. This is one of my most favourite movies of 2019”	['negative', 'reviews', 'one', 'favourite', 'movies']

favourite movies of 2019”	
---------------------------	--

5. Stemming

Tabel 7. Proses Stemming

Ulasan Awal	Sesudah Stemming
“A 3 hours Marvel film with 2.5 hours of drama, 15 mins of actions and 15 mins of something else.. seriously?Their logic didn't make any sense.Besides, what they did to Thor was brutal.”	['hour', 'marvel', 'film', 'hour', 'drama', 'min', 'action', 'min', 'someth', 'els', 'seriouslytheir', 'logic', 'didnt', 'make', 'sensebesid', 'thor', 'brutal']
“I LOVE MARVEL!!! This movie is incredible!! I just wish Stan lee was here to see everything he had worked to create!!”	['love', 'marvel', 'movi', 'incred', 'wish', 'stan', 'lee', 'see', 'everyth', 'work', 'creat']
“No Negative reviews.. This is one of my most favourite movies of 2019”	['neg', 'review', 'one', 'favourit', 'movi']

3.3. Pembobotan Kata

Jumlah kata atau fitur dalam dokumen adalah salah satu cara *count vectorizer* menghitung kemunculan. Jumlah kemunculan sebuah kata atau fitur dalam setiap dokumen dihitung. Sebagai contoh, Tabel 8 menyajikan dokumen dan perhitungan algoritma *count vectorizer*.

Tabel 8. Hasil algoritma Count Vectorizer

Kosakata	Dok. 1	Dok. 2	Dok. 3
action	1	0	0
brutal	1	0	0
creat	0	1	0
didn't	1	0	0
drama	1	0	0
els	1	0	0
everyth	0	1	0
favourit	0	0	1
film	1	0	0

hour	2	0	0
incred	0	1	0
lee	0	1	0
logic	1	0	0
love	0	1	0
make	1	0	0
marvel	1	1	0
min	2	0	0
movi	0	1	1
neg	0	0	1
one	0	0	1
review	0	0	1
see	0	1	0
sensebesid	1	0	0
seriouslytheir	1	0	0
someth	1	0	0
stan	0	1	0
thor	1	0	0
wish	0	1	0
work	0	1	0

3.4. Klasifikasi Sentimen

Setelah 1500 data ulasan film “Avengers: Endgame” diproses dan diberi bobot menggunakan metode *count vectorizer*, langkah selanjutnya adalah melakukan klasifikasi sentimen dengan menggunakan dua algoritma berbeda. Kedua algoritma ini diterapkan untuk membagi ulasan menjadi dua kategori utama yaitu sentimen positif dan negatif.

Untuk memastikan model dapat menangani data baru dengan baik, dataset dibagi menjadi data latih sebanyak 20% dan data uji sebanyak 80%. Pembagian ini bertujuan agar model dapat mempelajari pola dari data latih dan diuji dengan data yang belum pernah dilihat sebelumnya, sehingga evaluasi performanya lebih objektif.

1. Klasifikasi Naïve Bayes

Classification report untuk model Naïve Bayes ditunjukkan pada Tabel 9.

Tabel 9. Naïve Bayes Classification Report

	Precision	Recall	F1-Score	Support
Negative	0.82	0.73	0.77	126
Positive	0.82	0.89	0.85	174

Accuracy				0.82	300
Macro Avg	0.82	0.81	0.81		300
Weighted Avg	0.82	0.82	0.82		300

Model Naïve Bayes menunjukkan performa yang baik dengan akurasi sebesar 82%.

2. Klasifikasi SVM

Selanjutnya, *classification report* untuk model SVM dengan menerapkan dataset yang sama. Hasilnya disajikan pada Tabel 10.

Tabel 10. SVM Classification Report

	Precision	Recall	F1-Score	Support
Negative	0.81	0.77	0.79	126
Positive	0.84	0.87	0.86	174

Accuracy				0.83	300
Macro Avg	0.83	0.82	0.82		300
Weighted Avg	0.83	0.83	0.83		300

Model Naïve Bayes menunjukkan performa yang lebih unggul dengan akurasi sebesar 83%.

3.5. Hasil Evaluasi

Pada tahap evaluasi, performa model klasifikasi Naïve Bayes dan SVM dalam mengklasifikasikan sentimen ulasan film “Avengers: Endgame” di IMDb dinilai menggunakan beberapa metrik, termasuk *confusion matrix* dan *word cloud*.

2. Word Cloud SVM

Gambar 7. Word Cloud Positif SVM



Gambar 7 menunjukkan hasil *word cloud* sentimen positif dari SVM, dimana kata yang paling sering muncul adalah “movie”, “amazing”, “great”, “good”, dan “best”.

Gambar 8. Word Cloud Negatif SVM



Gambar 8 menunjukkan hasil *word cloud* sentimen negatif dari SVM, dimana kata yang paling sering muncul adalah “movie”, “boring”, “long”, “Infinity War”, dan “Marvel”.

4. PENUTUP

4.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan mengenai komparasi analisis sentimen ulasan film “*Avengers: Endgame*” di IMDB menggunakan metode Naïve Bayes dan SVM yang menggunakan 1500 dataset, dapat disimpulkan bahwa algoritma SVM sedikit lebih unggul dibandingkan Naïve Bayes dengan

akurasi SVM sebesar 83% dan Naïve Bayes sebesar 82%.

Dari evaluasi hasil menggunakan *confusion matrix* dan *classification report* pada algoritma SVM dapat disimpulkan bahwa ulasan film “*Avengers: Endgame*” tergolong positif dengan nilai *precision* 84%, *recall* 87%, dan *f1-score* 86%. Nilai tersebut mencerminkan bahwa model SVM tidak hanya akurat dalam mengidentifikasi ulasan positif tetapi juga konsisten dalam mengenali pola sentimen yang ada. Meskipun demikian, model Naïve Bayes juga menunjukkan kinerja yang cukup baik, meskipun sedikit tertinggal dibandingkan SVM.

Hasil dari kedua model ini memberikan gambaran yang lebih jelas tentang bagaimana penonton merespons film “*Avengers: Endgame*” di platform IMDB, dimana sentimen positif lebih dominan. Penelitian ini menunjukkan bahwa penggunaan teknik pemrosesan bahasa alami dengan algoritma yang tepat dapat memberikan wawasan yang berharga tentang pandangan publik terhadap sebuah film.

4.2. Saran

- Disarankan untuk menggunakan dataset yang lebih besar dan lebih beragam, termasuk ulasan dari platform lain untuk meningkatkan generalisasi hasil dan validitas model.
- Eksplorasi algoritma lain seperti Random Forest atau Deep Learning dapat dilakukan untuk membandingkan performa klasifikasi sentimen, sehingga dapat ditemukan model yang lebih optimal.
- Penggunaan teknik pembobotan kata yang lebih kompleks, seperti word embeddings dapat dipertimbangkan untuk menangkap makna kata dalam konteks yang lebih baik.
- Integrasi analisis sentimen dengan analisis aspek atau emosi dapat memberikan wawasan yang lebih mendalam tentang sentimen yang diekspresikan dalam ulasan, yang tidak hanya terbatas pada kategori positif dan negatif.

5. DAFTAR PUSTAKA

- Agustin, R.I. (2024) 'KOMPARASI ALGORITMA NAÏVE BAYES DAN SVM UNTUK ANALISIS SENTIMEN TWITTER KORUPSI BANSOS BERAS MASA PANDEMI', *Jurnal Informatika dan Teknik Elektro Terapan*, 12(2). Available at: <https://doi.org/10.23960/jitet.v12i2.4020>
- Ardana, Y. and Kambau, R.A. (2023) 'PERBANDINGAN ANALISIS SENTIMEN ALGORITMA SUPPORT VECTOR MACHINE DAN NAÏVE BAYES TERHADAP TANGGAPAN PUBLIK TENTANG PEMBELAJARAN ONLINE DI MASA PANDEMI COVID-19', *JOURNAL OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE*, 3(1), pp. 1–7.
- Arif Widiawan Subagio, Anggraini Puspita Sari and Andreas Nugroho Sihananto (2024) 'Klasifikasi Lexicon-Based Sentiment Analysis Tragedi Kanjuruhan pada Twitter Menggunakan Algoritma Convolutional Neural Network', *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*, 4(1), pp. 166–177. Available at: <https://doi.org/10.55606/juisik.v4i1.759>.
- Assegaff, S., Rasywir, E. and Pratama, Y. (2023) 'Experimental of vectorizer and classifier for scrapped social media data', *Telkomnika (Telecommunication Computing Electronics and Control)*, 21(4), pp. 815–824. Available at: <https://doi.org/10.12928/TELKOMNIK A.v21i4.24180>.
- Chandani, V. and Wahono, R.S. (2015) 'Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film', *Journal of Intelligent Systems*, 1(1). Available at: <http://journal.ilmukomputer.org>.
- Darmawan, G. et al. (2023) 'ANALISIS SENTIMEN BERDASARKAN ULASAN PENGGUNA APLIKASI MYPERTAMINA PADA GOOGLE PLAYSTORE MENGGUNAKAN METODE NAÏVE BAYES INFO ARTIKEL ABSTRAK', 2(3), pp. 100–108. Available at: <https://doi.org/10.55123>.
- Darwis, D. et al. (2020) *PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA*, *Jurnal Ilmiah Edutic*.
- Frando Febriant, Hanna Christy and Andri Wijaya (2023) 'Analisis Sentimen Film The Marvels Dari Aplikasi Twitter Menggunakan Metode Lexicon Based', *Jurnal Sistem & Teknologi Informasi Komunikasi*, 7(1), pp. 37–42.
- Gracezando Yehova, P. and Ihsan Jambak, M. (2024) 'PERBANDINGAN METODE NAÏVE BAYES DAN SVM PADA ULASAN GOOGLE PLAYSTORE MOBILE LEGENDS: BANG BANG', *JURNAL DEVICE*, 14(1), pp. 140–149.
- Gunawan, P.H., Alhafidh, T.D. and Wahyudi, B.A. (2022) 'The Sentiment Analysis of Spider-Man: No Way Home Film Based on IMDb Reviews', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(1), pp. 177–182. Available at: <https://doi.org/10.29207/resti.v6i1.3851>.
- Noviana, R. and Rasal B A Jurusan, I. (2023) 'PENERAPAN ALGORITMA NAIVE BAYES DAN SVM UNTUK ANALISIS SENTIMEN BOY BAND BTS PADA MEDIA SOSIAL TWITTER', *JTS*, 2(2), pp. 51–60.
- Nurian, A. et al. (2024) 'ANALISIS SENTIMEN PENGGUNA APLIKASI SHOPEE PADA SITUS GOOGLE PLAY MENGGUNAKAN NAIVE BAYES CLASSIFIER', *Jurnal Informatika dan Teknik Elektro Terapan*, 12(1). Available at: <https://doi.org/10.23960/jitet.v12i1.3631>
- Nurian, A., Padilah, T.N. and Garno, G. (2024) 'ANALISIS SENTIMEN TERHADAP PELAYANAN DISDUKCAPIL KARAWANG MENGGUNAKAN NAIVE BAYES CLASSIFIER', *Jurnal Informatika dan Teknik Elektro Terapan*, 12(2). Available at: <https://doi.org/10.23960/jitet.v12i2.4178>

- Rahman, I.F., Hasanah, A.N. and Heryana, N. (2024) ‘ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SAMSAT DIGIITAL NASIONAL (SIGNAL) DENGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER’, *Jurnal Informatika dan Teknik Elektro Terapan*, 12(2). Available at: <https://doi.org/10.23960/jitet.v12i2.4073>
- Rifka Pratama (2022) ‘FILM DOCTOR STRANGE IN THE MULTIVERSE OF MADNESS (2022) SEBAGAI BUDAYA POPULER: SEBUAH PERSPEKTIF PENONTON INDONESIA’, *Endogami: Jurnal Ilmiah Kajian Antropologi*, 5(2), pp. 84–93.
- Sajid Khan *et al.* (2023) ‘Fake News Classification using Machine Learning: Count Vectorizer and Support Vector Machine’, *Journal of Computing & Biomedical Informatics*, 4(01). Available at: <https://doi.org/10.56979/401/2022/85>.