

Analisis Sentimen Pada Ulasan IMDb Dengan Algoritma Naïve Bayes dan SVM

Andri Tri Agustama*¹, Rina Firliana², Arie Nugroho³

¹Program Studi Sistem Informasi, Universitas Nusantara PGRI Kediri

E-mail: *¹mawan6989@gmail.com, ²rina@unpkediri.ac.id, ³arienugroho@unpkediri.ac.id

Abstrak

IMDb merupakan platform populer yang menyediakan ulasan film dari berbagai pengguna di seluruh dunia. Banyaknya ulasan yang tersedia menjadikan proses analisis secara manual tidak efisien dan memerlukan metode otomatis. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma klasifikasi teks, yaitu Naïve Bayes dan Support Vector Machine (SVM), dalam melakukan analisis sentimen terhadap ulasan film di IMDb. Proses analisis dimulai dengan tahapan preprocessing data, seperti pembersihan teks, tokenisasi, penghapusan stopword, dan stemming. Setelah itu, fitur teks diolah menggunakan teknik pembobotan TF-IDF untuk menghasilkan representasi numerik. Dataset kemudian dibagi menjadi data latih dan data uji untuk mengukur performa model menggunakan metrik seperti akurasi, presisi, recall, dan f1-score. Hasil evaluasi menunjukkan bahwa algoritma SVM memperoleh akurasi sebesar 89%, sementara Naïve Bayes mencapai akurasi 87%. Perbedaan ini menunjukkan bahwa SVM memiliki keunggulan dalam klasifikasi yang lebih akurat, sedangkan Naïve Bayes tetap kompetitif dalam hal kecepatan dan efisiensi. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem analisis sentimen otomatis yang lebih akurat dan efisien, khususnya untuk ulasan berbasis teks seperti pada IMDb.

Kata Kunci—Analisis Sentimen, IMDb, Naïve bayes, support Vector Machine

Abstract

IMDb is a popular platform that provides film reviews from users around the world. The vast number of reviews makes manual analysis inefficient, highlighting the need for automated sentiment classification methods. This study aims to compare the performance of two text classification algorithms—Naïve Bayes and Support Vector Machine (SVM)—in sentiment analysis of IMDb movie reviews. The research begins with several preprocessing steps, including text cleaning, tokenization, stopword removal, and stemming. Next, the TF-IDF weighting technique is applied to convert textual data into numerical features. The dataset is split into training and testing sets to evaluate model performance using metrics such as accuracy, precision, recall, and F1-score. Evaluation results show that the SVM algorithm achieved an accuracy of 89%, while Naïve Bayes reached 87%. These results indicate that while SVM provides higher classification accuracy, Naïve Bayes remains competitive in terms of processing speed and efficiency. Therefore, the findings of this research are expected to contribute to the development of more accurate and efficient sentiment analysis systems, particularly for text-based reviews like those found on IMDb.

Keywords—Sentiment Analisis, Naïve bayes, Support vector machine

1. PENDAHULUAN

IMDb merupakan salah satu platform daring terkemuka yang menyediakan berbagai informasi terkait film dan acara televisi. Salah satu fitur utama dari platform ini adalah kemampuan pengguna untuk memberikan penilaian dan ulasan, sehingga membentuk basis data opini publik yang sangat besar dan beragam [1]. Ulasan-ulasan tersebut mencerminkan pandangan masyarakat terhadap suatu karya dan

memberikan referensi tambahan bagi calon penonton dalam menilai kualitas sebuah film [2]. Namun, volume ulasan yang sangat besar membuat proses analisis secara manual menjadi tidak efisien dan memakan waktu, sehingga dibutuhkan pendekatan otomatis untuk memproses data secara efektif [3].

Tantangan utama dalam analisis sentimen berasal dari bentuk ulasan yang tidak terstruktur, keberagaman bahasa, penggunaan slang, emotikon, dan istilah modern, yang dapat menghambat pemahaman serta klasifikasi sentimen[4]. Berbagai penelitian terdahulu telah memanfaatkan algoritma seperti Support Vector Machine (SVM) dan Naïve Bayes yang dikombinasikan dengan metode pembobotan TF-IDF. SVM dikenal mampu mengelola data kompleks berdimensi tinggi, sedangkan Naïve Bayes unggul dalam kecepatan dan kesederhanaannya[5]. Tahapan preprocessing seperti tokenisasi, stemming, dan penghapusan stopword juga terbukti meningkatkan akurasi klasifikasi[6].

Penerapan analisis sentimen secara otomatis pada ulasan IMDb tidak hanya memudahkan pengguna dalam memilih film, tetapi juga memberikan masukan penting bagi pembuat film dalam memahami respons penonton[7]. Seiring meningkatnya ketergantungan pada ulasan digital dalam pengambilan keputusan, penelitian ini menjadi semakin relevan[8]. Dengan menggabungkan metode TF-IDF dan algoritma SVM serta Naïve Bayes, penelitian ini bertujuan untuk mengatasi tantangan dalam klasifikasi sentimen dan menentukan pendekatan yang paling efektif. Berdasarkan studi sebelumnya, meskipun keduanya menunjukkan performa yang baik, hasil evaluasi menunjukkan variasi tingkat akurasi, seperti pada Twitter review di mana Naïve Bayes mencapai akurasi 84% dan SVM hingga 98%, namun disimpulkan Naïve Bayes lebih unggul[9].

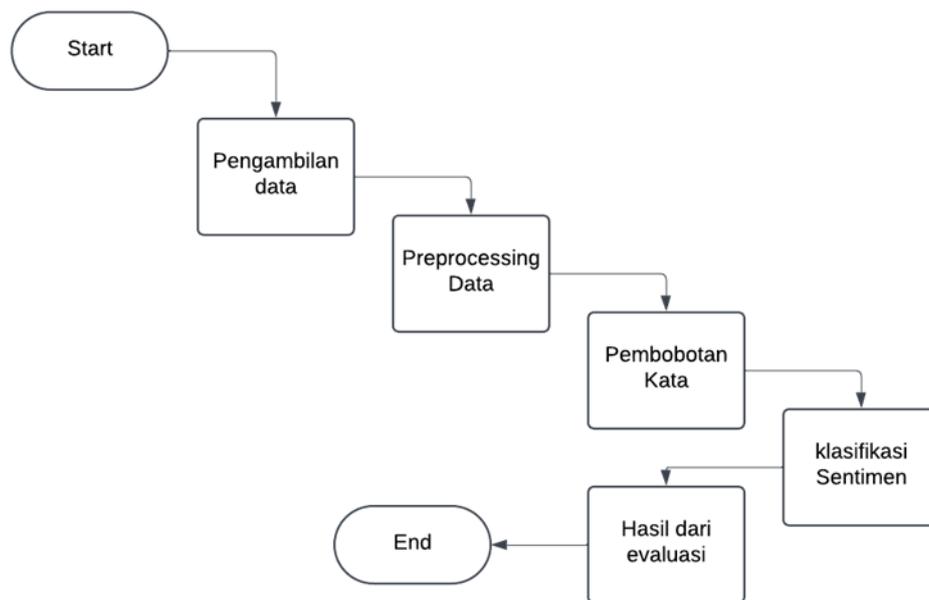
Penelitian ini bertujuan untuk membandingkan performa SVM dan Naïve Bayes dalam klasifikasi sentimen ulasan film IMDb dengan dukungan TF-IDF dan preprocessing teks, guna mengidentifikasi metode yang paling optimal. Hasil akhir diharapkan dapat memberikan kontribusi pada pengembangan sistem analisis sentimen berbasis teks, baik untuk penerapan praktis maupun penelitian lanjutan dalam text mining[10].

2. METODE PENELITIAN

Penelitian ini menggunakan metode *Machine Learning* salah satu pendekatan yang umum digunakan dalam analisis sentimen karena kemampuannya dalam mengenali pola dari data historis yang sudah diberi label. Pendekatan ini memanfaatkan model klasifikasi yang dilatih menggunakan dataset dengan anotasi sentimen[11], seperti positif, negatif. Sebelum proses pelatihan dilakukan, data teks biasanya melalui tahap praproses untuk membersihkan elemen-elemen yang tidak relevan, lalu dikonversi ke dalam bentuk numerik menggunakan teknik seperti *Term Frequency-Inverse Document Frequency (TF-IDF)*.

2.1. Alur penelitian

Pada gambar 1 adalah alur dari penelitian dimulai dengan tahap pengumpulan data, di mana data yang relevan dengan objek penelitian diambil dari sumber tertentu.



Gambar 1. Alur penelitian.

Tahap ini berperan penting untuk memastikan ketersediaan data mentah yang akan digunakan dalam proses analisis selanjutnya, data yang telah dikumpulkan masuk ke tahap preprocessing, yang mencakup proses pembersihan dan normalisasi data, seperti penghapusan data duplikat, penghilangan tanda baca, serta tokenisasi.

2.2. Sumber Data

Sumber data dalam studi ini dilakukan dengan memanfaatkan dataset yang tersedia secara publik melalui situs Kaggle. Dataset tersebut berisi ulasan-ulasan film dari platform *IMDb* yang telah diklasifikasikan ke dalam dua jenis sentimen, yaitu positif dan negatif. Pemilihan sumber data ini didasarkan pada kemudahan akses, kelengkapan informasi, serta format data yang sudah tersusun secara sistematis dan sesuai untuk keperluan analisis sentimen berbasis teks. Pendekatan ini juga dinilai lebih efisien karena tidak memerlukan proses pengambilan data langsung dari situs *IMDb* secara manual.

2.3. Tahap pre processing

Tahap preprocessing dalam analisis sentimen mencakup pembersihan data teks melalui penghapusan tanda baca, angka, dan kata-kata tidak penting, serta proses normalisasi seperti case folding, stopwords removal, stemming, dan tokenisasi. Data yang telah dibersihkan kemudian diubah menjadi representasi numerik menggunakan metode TF-IDF untuk memberi bobot pentingnya kata dalam dokumen. Hasilnya digunakan sebagai input bagi model klasifikasi sentimen, *Naïve Bayes* dan *SVM*[12]. Kinerja model kemudian dievaluasi menggunakan metrik seperti akurasi untuk memastikan metode yang digunakan sesuai dengan tujuan penelitian[13].

2.4. Pembobotan Kata

Setelah tahap preprocessing selesai, data teks diproses lebih lanjut melalui pembobotan kata menggunakan teknik seperti Term Frequency-Inverse Document Frequency (TF-IDF) untuk menghitung bobot numerik setiap kata berdasarkan frekuensi kemunculannya dalam dokumen tertentu dan tingkat kelangkaannya di seluruh dokumen, sehingga menghasilkan representasi vektor numerik yang mencerminkan pentingnya kata tersebut dalam konteks analisis. Representasi ini kemudian digunakan dalam tahap klasifikasi sentimen dengan algoritma pembelajaran mesin tradisional seperti *Naïve Bayes* atau Support Vector Machine (SVM) yang dilatih menggunakan data berlabel (positif,

negatif, atau netral) agar dapat mengenali pola sentimen dalam data baru secara otomatis. Tahap akhir, evaluasi model, dilakukan dengan metrik seperti akurasi, presisi, recall, dan F1-score untuk menilai sejauh mana metode yang diterapkan berhasil memenuhi tujuan penelitian.

3. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil yang dicapai dalam penelitian ini, yang meliputi tahap pengambilan data, proses pembersihan data, pembobotan kata, klasifikasi sentimen, dan evaluasi performa model [14]. Data yang digunakan telah melewati proses pra-pemrosesan secara menyeluruh agar lebih siap untuk dianalisis lebih lanjut. Selanjutnya, pembobotan kata dilakukan dengan teknik *TF-IDF* yang mengubah data teks menjadi bentuk representasi numerik. Setelah itu, data yang telah diolah digunakan dalam tahap klasifikasi dengan menggunakan dua algoritma utama, yakni Naïve Bayes dan *Support Vector Machine (SVM)*. Performa kedua model kemudian dievaluasi menggunakan metrik seperti *akurasi, presisi, recall, dan f1-score* untuk melihat efektivitasnya secara keseluruhan.

3.1. Data IMDb

Data ulasan berhasil diambil dari situs web IMDb mengambil dari situs web Kaggle dengan labelling negative (1) positif (0). Penggunaan situs web ini memungkinkan pengambilan data sejumlah besar ulasan dengan efisien dan akurat. Dataset dengan jumlah 50.000 yang terambil kemudian disimpan untuk tahap preprocessing selanjutnya. Contoh hasil tersedia dalam Tabel 1.

Tabel 1. Data IMDb

Index	Review	Sentiment
0	One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked. They are right, as this is exactly what happened with me...	positive
1	A wonderful little production. The filming technique is very unassuming...	positive
2	I thought this was a wonderful way to spend time on a too hot summer weekend...	positive
3	Basically there's a family where a little boy (Jake) thinks there's a zombie in his closet...	negative
4	Petter Mattei's 'Love in the Time of Money' is a visually stunning film to watch...	positive

3.2. Pre processing

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 Terdapat 5 tahapan yaitu cleaning text, Stopword removal, tokenize, stemming, pembobotan kata

3.3. Cleaning text

Pada table 4.1 adalah clean teks digunakan untuk Menghilangkan karakter-karakter yang tidak diperlukan seperti tanda baca, HTML tag, angka, dan simbol. Tujuannya adalah menyederhanakan isi teks agar lebih terstruktur.

Tabel 2. Cleaning text

No	Review	Cleaned Review
0	one of the other### review has mention that -after...	one of the other reviews has mention that after...
1	a wonder littl production. the fil @	a wonder little production. the fil
2	i thought this was a wonder way to spend time ()	i thought this was a wonder way to spend time
3	basic there a famili where a little boy (jake) ...	basic there a famili where a little boy jake
4	petter mattei "love in the time of money" is a...	petter mattei love in the time of money is a

Tabel tersebut menampilkan dua kolom utama, yaitu Review dan Cleaned Review, yang merupakan bagian dari proses preprocessing data dalam analisis sentimen berbasis teks. Tujuan utama dari tahap ini adalah membersihkan data teks agar lebih terstruktur dan relevan untuk dianalisis oleh model pembelajaran mesin atau algoritma pemrosesan bahasa alami (Natural Language Processing/NLP).

3.3.1 Stopword Removal

Pada table 3 adalah Stopword kata-kata umum yang muncul sangat sering namun tidak memiliki kontribusi penting dalam analisis makna, seperti: the, is, at, and, in, on, of, dll. Penghapusan stopwords dilakukan agar model tidak terganggu oleh kata-kata yang frekuensinya tinggi namun tidak membedakan konteks sentiment, dalam penelitian ini, digunakan daftar stopwords dari pustaka NLTK.

Tabel 3. Stopword Removal

sebelum	Sesudah stopwords
"One of the other reviewers has mentioned that after watching..."	['one', 'reviewers', 'has', 'mentioned', 'after', 'watching', 'oz', 'episode']
"A wonderful little production. The filming technique..."	['wonderful', 'little', 'production', 'br', 'br', 'filming', 'technique', 'unassuming']
"I thought this was a wonderful way to spend time..."	['i', 'thought', 'wonderful', 'way', 'spend', 'time', 'hot', 'summer', 'afternoon']
"Basically there's a family where a little boy Jake..."	['basically', 'theres', 'family', 'little', 'boy', 'jake', 'thinks', 'theres', 'zombie', 'closet']

Setiap baris mencerminkan satu entri ulasan yang telah diproses untuk menghapus stopwords, yakni kata-kata umum yang tidak memberikan kontribusi besar terhadap makna (seperti "the", "is", "and", dan sebagainya).

3.3.2 Tokenization

Pada table 4 adalah proses memecah teks ulasan menjadi unit-unit kata (token).

Tabel 4. Tokenization.

Index	Tokenized Text
0	One, of, the, other, reviewers, has, mentioned
1	A, wonderful, little, production.
2	I, thought, this, was, a, wonderful, way, to
3	Basically, there's, a, family, where, a, litt
4	Petter, Mattei's, "Love, in, the, Time, of, M

Tokenisasi merupakan salah satu tahapan awal yang esensial dalam pemrosesan bahasa alami (Natural Language Processing). Proses ini bertujuan untuk memecah teks atau kalimat menjadi unit-unit linguistik yang lebih kecil, yang dikenal sebagai token. Token dapat berupa kata, frasa, atau simbol tertentu tergantung pada pendekatan yang digunakan. Dalam konteks analisis teks ulasan film, seperti yang terdapat dalam dataset ini, tokenisasi berfungsi untuk mengubah setiap kalimat menjadi deretan kata yang terpisah. Misalnya, kalimat "I thought this was a wonderful way to spend time" akan diubah menjadi daftar token: ['I', 'thought', 'this', 'was', 'a', 'wonderful', 'way', 'to', 'spend', 'time']. Transformasi ini memungkinkan sistem komputasi untuk melakukan analisis lebih lanjut, seperti penghitungan frekuensi kata, klasifikasi sentimen, atau ekstraksi fitur secara otomatis.

3.3.3 Stemming

Pada tabel 5. stemming proses untuk mengembalikan kata ke bentuk dasarnya atau akar kata. Tujuannya untuk menyamakan kata yang memiliki arti sama tetapi bentuk berbeda.

Tabel 5. Stemming

Kalimat Asli (Ringkas)	Setelah Stemming
"One of the other reviewers has mentioned that after watching..."	['one', 'review', 'ha', 'mention', 'after', 'watch', 'oz', 'episod', 'hook']
"A wonderful little production. The filming technique..."	['wonder', 'littl', 'product', 'film', 'techniqu', 'veri', 'unassum', 'fashion', 'realism']
"I thought this was a wonderful way to spend time..."	['thought', 'wa', 'wonder', 'way', 'spend', 'time', 'hot', 'summer', 'weekend', 'comedi']
"Basically there's a family where a little boy Jake..."	['basic', 'famili', 'littl', 'boy', 'jake', 'think', 'zombi', 'closet', 'parent', 'fight']

Proses *stemming* yang saya terapkan di sini menghasilkan transformasi kata-kata dalam kalimat asli menjadi bentuk yang lebih ringkas, tanpa mempertimbangkan konteks gramatikal atau sintaktis. Meskipun hasil *stemming* tidak selalu menghasilkan kata yang sesuai secara

3.3.4 Pembobotan kata

Metode dari tabel 6 bekerja dengan dua komponen utama. Pertama, *Term Frequency (TF)* menghitung seberapa sering suatu kata muncul dalam satu dokumen.

Tabel 6. Tf-idf pembobotan kata.

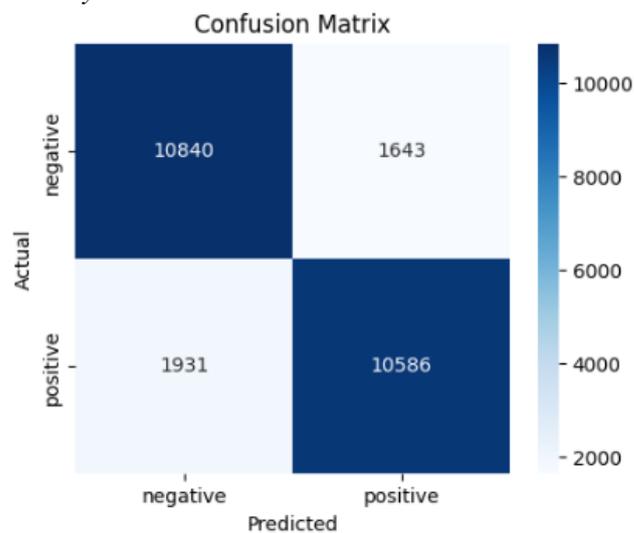
Kosakata	Dok.1	Dok.2	Dok.3
accustomed	1	0	0
actors	0	1	0
addiction	0	0	1
agenda	1	0	0
agreements	1	0	0

Semakin sering kata tersebut muncul, semakin besar nilai TF-nya. Kedua, *Inverse Document Frequency (IDF)* berfungsi untuk menurunkan bobot kata-kata yang umum dan sering muncul di banyak dokumen, karena kata-kata tersebut dianggap kurang informatif. Dengan mengalikan kedua komponen tersebut.

3.4. Hasil Evaluasi

Hasil Evaluasi Pada tahap evaluasi, performa model klasifikasi Naïve Bayes dalam mengklasifikasikan sentimen ulasan di IMDb dinilai menggunakan beberapa metrik, termasuk *confusion matrix* pada algoritma *naïve bayes*, *svm* dan *wordcloud*.

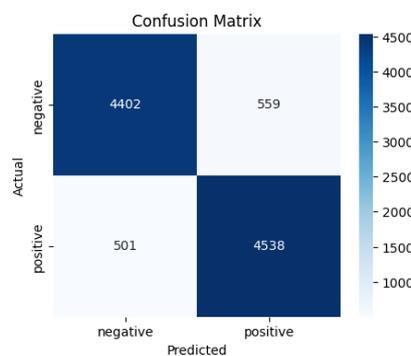
Confusion matrix Naïve bayes



Gambar 2. Confusion matrix

Gambar 2 merupakan *confusion matrix* pada algoritma Naïve bayes yang menggambarkan performa model klasifikasi dalam membedakan antara kelas negatif dan positif. Berdasarkan hasil yang ditampilkan, model berhasil mengklasifikasikan 10.840 data aktual negatif dengan benar sebagai negatif (*True Negative*), dan 10.586 data aktual positif dengan benar sebagai positif (*True Positive*). Namun, terdapat pula kesalahan klasifikasi, yaitu sebanyak 1.643 data negatif yang salah diprediksi sebagai positif (*False Positive*) dan 1.931 data positif yang salah diprediksi sebagai negatif (*False Negative*). Informasi ini penting dalam analisis kinerja model karena dapat digunakan untuk menghitung berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score, yang memberikan gambaran lebih menyeluruh mengenai kemampuan model dalam melakukan klasifikasi secara tepat.

Confusion matrix SVM



Gambar 3. Confusion Matrix SVM

Confusion matrix yang ditampilkan pada Gambar 3 menunjukkan hasil evaluasi model klasifikasi biner dalam bentuk *confusion matrix SVM*, yang digunakan untuk menilai kinerja model dalam membedakan antara kelas negatif dan positif. Berdasarkan gambar, terdapat 4.402 data aktual negatif yang berhasil diklasifikasikan dengan benar sebagai negatif (*True Negative*), dan 4.538 data aktual positif yang berhasil diklasifikasikan dengan benar sebagai positif (*True Positive*). Namun, model juga melakukan kesalahan, yaitu mengklasifikasikan 559 data negatif sebagai positif (*False Positive*), serta 501 data positif sebagai negatif (*False Negative*). Nilai-nilai ini dapat digunakan untuk menghitung metrik evaluasi lainnya seperti akurasi, presisi, recall, dan F1-score, yang penting untuk memahami seberapa baik model dalam mengidentifikasi.

Klasifikasi SVM

	precision	recall	f1-score	support
negative	0.90	0.89	0.89	4961
positive	0.89	0.90	0.90	5039
accuracy			0.89	10000
macro avg	0.89	0.89	0.89	10000
weighted avg	0.89	0.89	0.89	10000

Gambar 6. Akurasi

Gambar 6. tersebut merupakan hasil *classification report* yang menunjukkan kinerja model klasifikasi dalam hal *precision*, *recall*, dan *f1-score* untuk dua kelas: *negative* dan *positive*. Untuk kelas negatif, model memiliki nilai *precision* sebesar 0,90 dan *recall* sebesar 0,89, menghasilkan *f1-score* sebesar 0,89. Untuk kelas positif, *precision* tercatat sebesar 0,89 dan *recall* sebesar 0,90, dengan *f1-score* yang juga mencapai 0,90. Secara keseluruhan, nilai *accuracy* model mencapai 0,89 atau 89%, yang menunjukkan bahwa 89% dari total 10.000 data uji berhasil diklasifikasikan dengan benar. Nilai *macro average* dan *weighted average* untuk seluruh metrik juga konsisten pada angka 0,89, yang menandakan bahwa model mampu mempertahankan performa yang seimbang di antara kedua kelas seperti tabel perbandingan di bawah.

Tabel 7. Perbandingan Akurasi

Metrik	Naive Bayes	SVM
Precision (Negatif)	0.86	0.90
Recall (Negatif)	0.89	0.89
F1-Score (Negatif)	0.87	0.89
Precision (Positif)	0.88	0.89
Recall (Positif)	0.85	0.90
F1-Score (Positif)	0.87	0.90
Accuracy	0.87	0.89
Macro Average	0.87	0.89
Weighted Average	0.87	0.89

Jika dibandingkan dengan hasil sebelumnya yang menunjukkan akurasi sebesar 87%, maka model SVM memiliki performa yang lebih baik dengan akurasi 89%. Selain itu, *precision*, *recall*, dan *f1-score* pada masing-masing kelas juga menunjukkan peningkatan. Sebagai contoh, pada model sebelumnya *precision* kelas positif adalah 0,88, sedangkan pada model ini meningkat menjadi 0,89. Hal ini menunjukkan bahwa SVM memiliki kemampuan klasifikasi yang lebih tinggi dibandingkan dengan model Naive Bayes sebelumnya, baik dalam mengenali kelas positif maupun negatif. Maka, dapat disimpulkan bahwa SVM memiliki kinerja yang lebih unggul dan lebih seimbang dalam klasifikasi data.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen ulasan film di IMDb menggunakan dua algoritma klasifikasi, yaitu Naive Bayes dan Support Vector Machine (SVM), dengan dukungan teknik pembobotan TF-IDF dan tahapan preprocessing teks. Berdasarkan hasil yang diperoleh, dapat

disimpulkan bahwa proses preprocessing data memainkan peran penting dalam meningkatkan performa klasifikasi. Tahapan seperti cleaning, tokenisasi, stopword removal, dan stemming terbukti mampu mengubah data tidak terstruktur menjadi representasi yang lebih teratur dan bermakna. Teknik pembobotan TF-IDF juga memberikan kontribusi signifikan dengan merepresentasikan fitur teks secara numerik dan informatif.

Dari sisi performa model, algoritma Naïve Bayes mampu menghasilkan akurasi sebesar 87%, menunjukkan efisiensi dan kecepatan dalam pemrosesan data. Namun, algoritma SVM memberikan hasil yang sedikit lebih unggul dengan akurasi rata-rata sebesar 89%, menunjukkan kemampuannya dalam menangani data berdimensi tinggi dengan tingkat akurasi yang lebih stabil. Dengan demikian, dapat disimpulkan bahwa meskipun kedua algoritma bekerja dengan baik, SVM lebih direkomendasikan untuk kebutuhan yang mengutamakan akurasi, sedangkan Naïve Bayes cocok untuk aplikasi yang membutuhkan efisiensi waktu dan sumber daya komputasi. Secara keseluruhan, penelitian ini menunjukkan bahwa pemilihan algoritma perlu disesuaikan dengan konteks dan tujuan implementasi yang diinginkan.

5. SARAN

Berdasarkan hasil dan temuan dalam penelitian ini, terdapat beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut. Pertama, untuk meningkatkan akurasi dan pemahaman konteks yang lebih kompleks, disarankan agar penelitian selanjutnya mencoba algoritma lain seperti Random Forest, XGBoost, atau metode berbasis deep learning seperti LSTM dan BERT. Kedua, dataset yang digunakan dapat diperluas dengan memasukkan ulasan dari berbagai platform lain seperti Rotten Tomatoes atau Metacritic agar model memiliki kemampuan generalisasi yang lebih baik terhadap variasi gaya bahasa pengguna.

DAFTAR PUSTAKA

- [1] Darwis Alwan and M. A. Ridla, "Averaged Word2vec sebagai Ekstraksi Fitur pada Analisis Sentimen Ulasan Film di IMDb menggunakan Artificial Neural Network (ANN)," *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, vol. 9, no. 1, pp. 36–45, Feb. 2024, doi: 10.32528/justindo.v9i1.1204.
- [2] B. Noviansyah, M. Makmun Effendi, and Y. Achmad, "Sentiment Analysis of Oppenheimer Movie Reviews: Naïve Bayes Algorithm for Public Opinion," *Architecture and High Performance Computing*, vol. 6, no. 3, 2024, doi: 10.47709/cnipc.v6i3.4393.
- [3] D. D. Nur Cahyo *et al.*, "Sentiment Analysis for IMDb Movie Review Using Support Vector Machine (SVM) Method," *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 8, no. 2, pp. 90–95, Mar. 2023, doi: 10.25139/inform.v8i2.5700.
- [4] M. I. Wardah and S. D. Putra, "Implementasi Machine Learning Untuk Rekomendasi Film Di Imdb Menggunakan Collaborative Filtering Berdasarkan Analisa Sentimen IMDB," *Jurnal Manajemen Informatika Jayakarta*, vol. 2, no. 3, p. 243, Jul. 2022, doi: 10.52362/jmijayakarta.v2i3.868.
- [5] A. Z. Amrullah, A. Sofyan Anas, M. Adrian, and J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal*, vol. 2, no. 1, 2020, doi: 10.30812/bite.v2i1.804.
- [6] G. Cahyani, W. Widayani, S. D. Anggita, Y. Pristyanto, I. Ikmah, and A. Sidauruk, "Klasifikasi Data Review IMDb Berdasarkan Analisis Sentimen Menggunakan Algoritma Support Vector Machine," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1418, Jul. 2022, doi: 10.30865/mib.v6i3.4023.
- [7] W. Ningsih, B. Alfianda, R. Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 556–562, Feb. 2024, doi: 10.57152/malcom.v4i2.1253.

- [8] S. A. Helmayanti, F. Hamami, and R. Y. Fa'rifah, "PENERAPAN ALGORITMA TF-IDF DAN NAÏVE BAYES UNTUK ANALISIS SENTIMEN BERBASIS ASPEK ULASAN APLIKASI FLIP PADA GOOGLE PLAY STORE," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 1822–1834, Sep. 2023, doi: 10.35870/jimik.v4i3.415.
- [9] J. Friadi and D. E. Kurniawan, "Analisis Sentimen Ulasan Wisatawan Terhadap Alun-Alun Kota Batam: Perbandingan Kinerja Metode Naive Bayes dan Support Vector Machine," *Jurnal Sistem Informasi Bisnis*, vol. 14, no. 4, pp. 403–407, Oct. 2024, doi: 10.21456/vol14iss4pp403-407.
- [10] E. Daniati and H. Utama, "ANALISIS SENTIMEN DENGAN PENDEKATAN ENSEMBLE LEARNING DAN WORD EMBEDDING PADA TWITTER," 2023.
- [11] S. K. Dirjen *et al.*, "Terakreditasi SINTA Peringkat 2 Perbandingan Performansi Algoritma Pengklasifikasian Terpandu Untuk Kasus Penyakit Kardiovaskular," *masa berlaku mulai*, vol. 1, no. 3, pp. 998–1006, 2017.
- [12] T. Andriyanto and E. Daniati, "Classification in the Self Monitoring System for Chronic Kidney Failure Patients on Hemodialysis Therapy with SVM," *JINAV: Journal of Information and Visualization*, vol. 3, no. 2, pp. 131–140, Dec. 2022, doi: 10.35877/454ri.jinav1410.
- [13] T. Adriyanto, R. A. Ramadhani, R. Helilintar, and A. Ristyawan, "Classification of Dog and Cat Images using the CNN Method," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 203–208, Dec. 2022, doi: 10.33096/ilkom.v14i3.1116.203-208.
- [14] A. Ristyawan, A. Nugroho, and T. K. Amarya, "Optimasi Preprocessing Model Random Forest Untuk Prediksi Stroke," vol. 12, no. 1, pp. 29–44, 2025. DOI: <https://doi.org/10.35957/jatisi.v12i1.9587>