

Analisis Sentimen Komentar Film Merah Putih One For All Metode Naïve Bayes

Bintang Arya Widi Syahputra^{a,1,*}, Budi Hartono^{a,2}

^a Program Studi Teknik Informatika, Universitas Stikubank, Jl. Kendeng V Bendan Ngisor, Semarang 50233, Indonesia

¹ bintangarya0035@mhs.unisbank.ac.id *; ² budihartono@edu.unisbank.ac.id

* Korespondensi penulis

Submission:23/12/2025, Revision: 08/01/2026, Accepted : 08/01/2026

Abstract

Social media Twitter (X) has become a platform for expressing public opinion, including reactions to the animated film "Merah Putih: One For All." The film has garnered various criticisms regarding its graphic quality and the use of digital assets deemed unoriginal. Sentiment analysis of this public response is essential to provide an objective evaluation of society's reception of local animated works, identify specific aspects of audience concern, and offer valuable insights for the Indonesian animation industry to improve production quality. Automated sentiment classification using machine learning is an efficient solution for understanding patterns of public perception with large volumes of data, which is difficult to accomplish manually. This study classifies comments on the X platform into positive and negative sentiment categories using the Naïve Bayes method. Data collection was conducted by scraping posts from the account @tanyakanrl in August 2025, totaling 302 comments. The data underwent preprocessing stages, including cleaning, case folding, normalization, tokenizing, stopword removal, and stemming. Feature extraction utilized the TF-IDF (Term Frequency-Inverse Document Frequency) method to convert text data into numerical representation. Classification employed the Naïve Bayes algorithm with an 80:20 data split ratio for training and testing. Evaluation results indicate that the model achieved an Accuracy of 82%, Precision of 79%, Recall of 82%, and an F1-Score of 80%.

Keywords: Animated Film, Naïve Bayes, Sentiment Analysis, Twitter (X)

Abstrak

Media sosial X (dahulu Twitter) telah menjadi ruang untuk mengekspresikan opini publik, termasuk tanggapan terhadap film animasi "Merah Putih: One For All". Film tersebut menuai berbagai kritik terkait kualitas grafik dan penggunaan aset digital yang dinilai tidak orisinal. Analisis sentimen terhadap respons publik sangat penting untuk memberikan evaluasi objektif mengenai penerimaan masyarakat terhadap karya animasi lokal, mengidentifikasi aspek yang menjadi perhatian audiens, serta memberikan wawasan (*insight*) berharga bagi industri animasi Indonesia. Klasifikasi sentimen otomatis menggunakan pembelajaran mesin (*machine learning*) menjadi solusi efisien untuk memahami pola persepsi publik pada data berskala besar. Penelitian ini mengklasifikasikan komentar pada platform X ke dalam kategori sentimen positif dan negatif menggunakan metode Naïve Bayes. Pengumpulan data dilakukan melalui teknik *scraping* pada akun @tanyakanrl pada Agustus 2025 dengan total 302 komentar. Data melalui tahapan pra-pemrosesan (*preprocessing*) yang meliputi *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*. Ekstraksi fitur menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik. Klasifikasi dilakukan dengan rasio pembagian data 80:20 untuk pelatihan (*training*) dan pengujian (*testing*). Hasil evaluasi menunjukkan bahwa model mencapai akurasi 82%, presisi 79%, *recall* 82%, dan *F1-score* 80%.

Kata kunci: Analisis Sentimen, Film Animasi, Naïve Bayes, Twitter (X)

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



1. Pendahuluan

Seiring dengan perkembangan teknologi dan komunikasi di era digital, internet telah membawa perubahan signifikan dalam kehidupan bermasyarakat, khususnya dalam cara berinteraksi dan berkomunikasi. Media sosial Twitter, yang kini telah berganti nama menjadi X, merupakan wadah bagi penyampaian opini publik. Platform X memiliki karakteristik penyebaran informasi yang cepat dan bersifat terbuka bagi publik [1]. Film animasi memiliki popularitas yang sangat tinggi, tidak hanya sebagai sarana hiburan, tetapi juga sebagai media komunikasi, edukasi, serta penyampaian informasi. Fenomena ini menunjukkan bahwa film animasi berperan penting dalam menyampaikan nilai-nilai budaya dan sosial, serta memberikan dampak signifikan terhadap industri hiburan secara global [2].

Film animasi juga sering digunakan sebagai media untuk menanamkan nilai kebangsaan. Salah satunya adalah film animasi "Merah Putih: One For All" yang dirilis pada tahun 2025. Film tersebut diproduksi oleh rumah produksi Perfiki Kreasindo dan disutradarai oleh Toto Soegriwo. Namun, karya ini menuai banyak kritik dari masyarakat Indonesia, terutama terkait kualitas grafik dan isu penggunaan aset digital yang dinilai tidak orisinal. Fenomena tersebut memicu berbagai komentar di platform X, baik berupa sentimen negatif yang mengkritik kualitas visual dan membandingkannya dengan film animasi terdahulu seperti "Jumbo" dan "Keluarga Somat", maupun respons positif yang memberikan dukungan serta saran konstruktif untuk meningkatkan kualitas produksi pada masa mendatang [3].

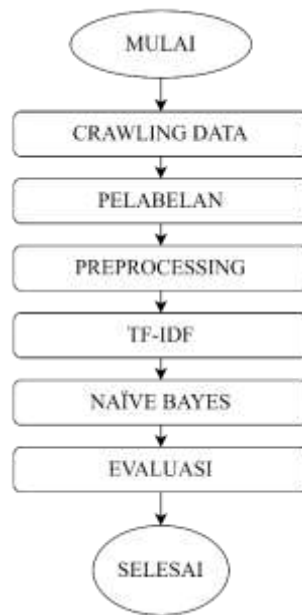
Publik menyoroti biaya produksi film yang mencapai Rp6,7 miliar dengan durasi pengerjaan selama dua bulan, yang dinilai tidak sebanding dengan kualitas hasil akhirnya. Komentar terhadap film animasi tersebut menunjukkan bahwa media sosial X berfungsi sebagai wadah bagi masyarakat untuk menyampaikan opini secara bebas. Platform X menjadi sarana yang paling efektif dalam mengekspresikan pandangan terhadap berbagai isu, termasuk pada film "Merah Putih: One For All". Terhadap karya tersebut, banyak pengguna memberikan tanggapan, baik sebelum maupun sesudah mereka menonton film tersebut [4].

Pendekatan yang digunakan untuk memahami secara mendalam tanggapan publik terkait film animasi tersebut adalah analisis sentimen. Analisis sentimen berfungsi untuk mengevaluasi opini, perasaan, dan emosi pengguna melalui data komentar berbentuk teks. Teknik ini mengidentifikasi komentar yang kemudian dikelompokkan ke dalam kategori sentimen positif dan negatif [5]. Pendekatan ini mengukur persepsi publik terhadap suatu fenomena yang terjadi, seperti pada film animasi Merah Putih: One For All [6].

Penambangan teks (*text mining*) merupakan salah satu teknik klasifikasi yang dapat menemukan pola dari sekumpulan data. Dalam penelitian ini, metode klasifikasi yang digunakan adalah Naïve Bayes [7]. Metode ini bekerja dengan menghitung probabilitas kemunculan kata dari masing-masing kelas sentimen yang berupa positif dan negatif. Menurut [8] Naïve Bayes merupakan algoritma yang umum digunakan dalam penelitian analisis sentimen karena memiliki tingkat akurasi tinggi dan sederhana dalam memproses data teks. Penelitian [2] tentang analisis sentimen terhadap sebuah film animasi dengan judul penelitian Analisis Sentimen Pengguna Twitter (X) Terkait Film One Piece Menggunakan Metode Algoritma Naive Bayes. Penelitian tersebut juga menerapkan pembobotan kata TF-IDF hingga mendapatkan tingkat akurasi 83%, dengan nilai *Precision* 77%, *Recall* 92%, *F1-Score* 84%. Dari penelitian ini menunjukkan bahwa metode ini memberikan hasil dengan cukup baik dalam mengklasifikasi sentimen di media sosial khususnya pada X.

2. Metode Penelitian

Metode penelitian diterapkan dengan melalui berbagai tahapan yang ditampilkan pada Gambar 1, mulai dari tahap *crawling* data, pelabelan, *preprocessing*, TF-IDF, Naïve Bayes, dan terakhir Evaluasi model.



Gambar 1. Tahapan Penelitian

2.1. Crawling Data

Dari Gambar 2, data yang dikumpulkan dan diambil dari platform X pada salah satu postingan akun “@tanyakanrl” dengan menggunakan *library tweet-harvest* sebagai *tools* untuk melakukan *crawling* data [9]. Data yang diambil dengan memakai link pada postingan tersebut seperti yang digunakan sebagai acuan pengambilan data, URL <https://x.com/tanyakanrl/status/1953427254463156479>, serta dengan periode pengambilan data tweet 7 – 31 Agustus.



Gambar 2. Postingan Media Sosial X

2.2. Pelabelan

Proses pelabelan data dilakukan secara manual untuk menentukan kategori pada setiap komentar yang telah dikumpulkan melalui tahap *crawling* data. Pelabelan sentimen ditentukan berdasarkan makna dan tujuan dari setiap kalimat serta kata yang terkandung di dalam teks. Selanjutnya, data sentimen tersebut dikelompokkan ke dalam kategori positif dan negatif.

2.3. Preprocessing

Preprocessing merupakan tahapan yang dilakukan untuk mengolah data teks dengan membersihkan teks dari link (URL) dan mengubah kata menjadi bentuk dasar [10]. Tahapan *preprocessing* meliputi, *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, *stemming*. Berikut penjelasan dari tahapan pada *preprocessing*.

- a. *Cleaning*
Tahap pembersihan (*cleaning*) merupakan proses penghapusan elemen yang tidak diperlukan dari sekumpulan data teks. Elemen-elemen tersebut meliputi tautan (URL), tanda baca, serta karakter lain yang tidak memiliki makna dalam proses analisis.
- b. *Case Folding*
Tahap *case folding* digunakan untuk menyeragamkan penulisan setiap kata di dalam kalimat sehingga data memiliki format yang konsisten saat diproses. Tahapan ini dilakukan dengan mengubah seluruh huruf kapital menjadi huruf kecil (*lowercase*). Sebagai contoh, frasa "Merah Putih" diubah menjadi "merah putih".
- c. Normalisasi
Normalisasi merupakan proses untuk merubah penulisan kata agar konsisten dan sesuai dengan kaidah bahasa yang baku, seperti "pak" menjadi "bapak" dan "bgt" menjadi "banget".
- d. *Tokenizing*
Tahapan *tokenizing* merupakan tahapan yang berfungsi membagi suatu dokumen teks menjadi unit-unit kata atau biasa disebut token, seperti pada kalimat "mungkin lagi bedah film" menjadi "mungkin", "lagi", "bedah", "film".
- e. *Stopword Removal*
Tahapan *Stopword Removal* adalah tahap untuk menghilangkan kata-kata umum dan sering muncul seperti "yang", "pada", "dan", atau "dari", karena kata-kata tersebut tidak memberikan makna yang signifikan terhadap isi kalimat.
- f. *Stemming*
Stemming adalah proses mengubah kata dalam teks ke bentuk dasarnya dengan menghapus imbuhan awalan, sisipan, atau akhiran, sehingga pada kata "tersirat" diubah menjadi "sirat" dan "membuat" menjadi "buat".

2.4. Pembobotan Kata TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan teknik ekstraksi fitur dari kata-kata didalam dokumen yang telah dipreprocessing sebelumnya, kemudian mengubah data teks menjadi bentuk numerik [11]. Nilai dari *Term Frequency* (TF) adalah teknik untuk menentukan jumlah kata yang muncul di dalam dokumen. Sedangkan nilai *Inverse Document Document* (IDF) bertujuan untuk mengukur suatu kata berdasarkan tingkat kemunculan di dalam seluruh dokumen. Mencari nilai dari TF-IDF adalah dengan mengkalikan nilai dari hasil *Term Frequency* (TF) dan nilai *Inverse Document Frequency* (IDF) [12].

$$tf - idf = tf * idf \quad (1)$$

2.5. Naïve Bayes

Naïve Bayes merupakan salah satu metode klasifikasi di dalam *text mining*, metode ini berguna saat menggunakan data teks yang berdimensi tinggi. Penerapan metode ini dapat bekerja dalam menghitung probabilitas dari suatu data berdasarkan frekuensi kemunculan kata dalam data.

$$P(C|X) = \frac{P(X|c) \times P(C)}{P(X)} \quad (2)$$

- X = Data dengan kelas yang belum diketahui
 C = Hipotesis data merupakan kelas spesifik
 P(C) = Probabilitas hipotesis C
 P(C|X) = Probabilitas hipotesis C berdasarkan kondisi X
 P(X) = Probabilitas hipotesis X
 P(X|C) = Probabilitas X berdasarkan kondisi C

2.6. Evaluasi Model

Proses perhitungan ditujukan untuk melihat hasil nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* yang diperoleh berdasarkan hasil model klasifikasi Naïve Bayes serta mengukur seberapa baik model bekerja. Perhitungan menggunakan empat komponen terdiri dari (TP) *True Positive*, (TN) *True Negative*, (FP) *False Positive*, (FN) *False Negative*.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1-Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{6}$$

3. Hasil dan Pembahasan

3.1. Crawling Data

Tahap crawling data dilakukan menggunakan teknik *crawling* Twitter (X) melalui tool *library tweet-harvest* menggunakan *twitter_auth_token* sebagai aksesn untuk memperoleh data berupa teks. Pada Gambar 3 adalah data hasil *crawling* yang disimpan dalam format file csv.

Tabel 1. Hasil Crawling Data

conversat	created_a	favorite_c	full_text	id_str	image_url	in_reply_t
1,95E+18	Thu Aug 0	140	@tanyakanrI posternya aja kalah sama bikin poster osis SMA wkwk ada canva p	1,95E+18		tanyakanr
1,95E+18	Fri Aug 08	125	@tanyakanrI jgn di tonton gaes. bentuk kita mendukung negara dengan support n	1,95E+18		tanyakanr
1,95E+18	Thu Aug 0	139	@tanyakanrI fomo bikin film animasi boleh sih. cuma tolonglah ceritanya itu dike	1,95E+18		tanyakanr
1,95E+18	Thu Aug 0	107	@tanyakanrI ini bikinnya pake Zepeto bukan???	1,95E+18		tanyakanr
1,95E+18	Thu Aug 0	91	@tanyakanrI Modal nasionalis ini karya anak bangsa no kritik positif selalu . Padal	1,95E+18		tanyakanr
1,95E+18	Thu Aug 0	95	@tanyakanrI https://t.co/Z2kw6rjXzd	1,95E+18	https://pt	tanyakanr
1,95E+18	Thu Aug 0	76	@tanyakanrI Nih trailernya	1,95E+18		tanyakanr
1,95E+18	Thu Aug 0	68	@tanyakanrI Pas banget ditanggal segitu nih animasi bisa bersaing dengan demor	1,95E+18		tanyakanr
1,95E+18	Thu Aug 0	55	@tanyakanrI ini yg pengerjaannya tiga bulan itu ya?	1,95E+18		tanyakanr
1,95E+18	Thu Aug 0	52	@tanyakanrI Jelek mana bocah baju ijo di tengah itu kek gibran lagi	1,95E+18		tanyakanr
1,95E+18	Thu Aug 0	52	@tanyakanrI Bertema kebangsaan judulnya pake bahasa Inggris https://t.co/GmV	1,95E+18	https://pt	tanyakanr
1,95E+18	Thu Aug 0	46	@tanyakanrI Posternya jelek font tidak cantik perpaduan warna kurang pas terlah	1,95E+18		tanyakanr
1,95E+18	Thu Aug 0	40	@tanyakanrI typography jelek colouring jelek alligment jelek jeleknya kok kompli	1,95E+18		tanyakanr
1,95E+18	Thu Aug 0	40	@tanyakanrI Kalo cuma tayang di TV mah oke2 aja. Malah kayaknya lebih gampan	1,95E+18		tanyakanr
1,95E+18	Fri Aug 08	36	@tanyakanrI Battle of Surabaya dianggap apa klo ini film animasi pertama anak in	1,95E+18	https://pt	tanyakanr

Proses *crawling* diketahui data yang peroleh sebanyak 302 *tweet* yang diambil dari satu postingan, pada ukuran dataset ini lebih kecil dibandingkan [13] yang menghasilkan jumlah data sebanyak 5000 dataset. Pada jumlah data yang diperoleh juga berdasarkan seberapa banyak komentar di dalam postingan tersebut. Dari data yang didapat kemudian diambil kolom “*full_text*” untuk selanjutnya dilakukan proses pelabelan sentimen positif dan negatif.

3.2. Pelabelan Data

Pelabelan data dilakukan secara manual dengan tujuan memastikan setiap *tweet* diklasifikasi sesuai dengan konteks dan makna kalimat [14]. Dari hasil pelabelan yang telah dilakukan menunjukkan bahwa dari 302 data *tweet*, terdapat 62 komentar dengan sentimen positif dan 240 komentar dengan sentimen negatif. Data pelabelan ditampilkan pada Tabel 1 berikut ini.

Tabel 2. Data Hasil Pelabelan

Full_text	Sentimen
@tanyakanrI posternya aja kalah sama bikin poster osis SMA wkwk ada canva prem loo yg ga terlalu mahal buat setingkat pemerintah mah tinggal gabungin element nya semisal emang MALES bgt ini mah keliatan kaya main main bgt	Positif
@tanyakanrI jgn di tonton gaes. bentuk kita mendukung negara dengan support mereka supaya bisa bikin film lebih baik. kalo penonton dikit mereka rugi pasti dari situ mikir apa yang salah? pls lah alur kayak gini tuh bagus buat bangun nasionalisme utk anak	Negatif
@tanyakanrI fomo bikin film animasi boleh sih. cuma tolonglah ceritanya itu dikemas yang menarik desain poster diperbaiki nanti kalo bisa ada subtitlenya ini tadi di trailer pas ada yang manggil nama itu kreskek kreskek	Negatif
@tanyakanrI Ini bikinnya pake Zepeto bukan???	Negatif
@tanyakanrI Modal nasionalis ini karya anak bangsa no kritik positif selalu . Padahal animasi luar aja malu kalo tau produknya belum siap tapi ditayangin makanya berjuang lebih biar bisa bagus tanpa berlingdung dibalik pernyataan karya anak bangsa bahkan d	Positif

3.3. Preprocessing

Setelah data dilakukan pelabelan secara manual, tahap selanjutnya adalah mengolah data dengan *preprocessing*. Pada penelitian [15] menerapkan 4 proses *preprocessing* seperti *case folding*, *cleaning*, *stopeord removal*, dan tokenisasi. sehingga tahap ini bertujuan untuk meningkatkan kualitas data teks dengan menambahkan proses tahap *preprocessing* seperti normalisasi dan *stemming* sebelum dilakukan pembobotan fitur dan klasifikasi. Sehingga proses ini mencakup tahapan *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming* [6]. Berbagai tahapan *preprocessing* tersebut ditampilkan pada Tabel 2.

Tabel 3. Hasil Tahapan Preprocessing

<i>Celaning</i>	<i>Case Folding</i>	Normalisasi	<i>Tokenizing</i>	<i>Stopword Removal</i>	<i>Stemming</i>
Posternya aja kalah sama bikin poster osis SMA wkwk ada canva prem loo yg ga terlalu mahal buat setingkat pemerintah mah tinggal gabungin element nya semisal emang MALES bgt ini mah keliatan kaya main main bgt	posternya aja kalah sama bikin poster osis sma wkwk ada canva prem loo yg ga terlalu mahal buat setingkat pemerintah mah tinggal gabungin element nya semisal emang males bgt ini mah keliatan kaya main main bgt	posternya saja kalah sama bikin poster osis sama wkwk ada canva prem lo yang tidak terlalu mahal buat setingkat pemerintah mah tinggal gabungin element ya semisal memang malas banget ini mah kelihatan kayak main main banget	['posternya', 'saja', 'kalah', 'sama', 'bikin', 'poster', 'osis', 'sama', 'wkwk', 'ada', 'canva', 'prem', 'lo', 'yang', 'tidak', 'yang', 'tidak', 'terlalu', 'mahal', 'buat', 'setingkat', 'pemerintah', 'mah', 'tinggal', 'gabungin', 'setingkat', 'pemerintah', 'mah', 'tinggal', 'gabungin', 'element', 'ya', 'semisal', 'memang', 'malas', 'banget', 'ini', 'mah', 'kelihatan', 'kayak', 'main', 'main', 'banget']	posternya kalah bikin poster osis wkwk canva prem mahal setingkat pemerintah tinggal gabungin element malas main main	poster kalah bikin poster osis wkwk canva prem mahal tingkat perintah tinggal gabungin element malas main main
Ini logo HUT RI nya gak ngikut logo pemerintah ya Judul nya pake enggres pulak ada adegan melawan uchiha Madara gak	ini logo hut ri nya gak ngikut logo pemerintah ya judul nya pake enggres pulak ada adegan melawan uchiha madara gak	ini logo hut ri ya tidak ngikut logo pemerintah ya judul ya pake enggres pula ada adegan melawan uchiha madara tidak	['ini', 'logo', 'hut', 'ri', 'ya', 'tidak', 'ngikut', 'logo', 'pemerintah', 'ya', 'judul', 'ya', 'pake', 'enggres', 'pula', 'ada', 'adegan', 'melawan', 'uchiha', 'madara', 'tidak']	logo hut ri ngikut logo pemerintah judul pake enggres adegan melawan uchiha madara	logo hut ri ngikut logo perintah judul pake enggres adegan lawan uchiha madara
Battle of Surabaya dianggap apa klo ini film animasi pertama bertema kebangsaan	battle of surabaya dianggap apa klo ini film animasi pertama anak indonesia bertema kebangsaan	battle of surabaya dianggap apa kalau ini film animasi pertama anak indonesia bertema kebangsaan	['battle', 'of', 'surabaya', 'dianggap', 'apa', 'kalau', 'ini', 'film', 'animasi', 'pertama', 'anak', 'indonesia', 'bertema', 'kebangsaan']	battle surabaya dianggap film animasi anak indonesia bertema kebangsaan	battle surabaya anggap film animasi anak indonesia tema bangsa
Jgn di tonton gaes bentuk kita mendukung negara dengan support mereka supaya bisa bikin film lebih baik kalo penonton dikit mereka rugi pasti dari situ mikir apa yang salah pls lah alur kayak gini tuh bagus buat bangun nasionalisme	jgn di tonton gaes bentuk kita mendukung negara dengan support mereka supaya bisa bikin film lebih baik kalo penonton dikit mereka rugi pasti dari situ mikir apa yang salah pls lah alur kayak gini tuh bagus buat bangun nasionalisme	jangan di tonton gaes bentuk kita mendukung negara dengan support mereka supaya bisa bikin film lebih baik kalau penonton dikit mereka rugi pasti dari situ mikir apa yang salah pls lah alur kayak gini tuh bagus buat bangun nasionalisme	['jangan', 'di', 'tonton', 'gaes', 'bentuk', 'kita', 'mendukung', 'negara', 'dengan', 'support', 'mereka', 'supaya', 'bisa', 'bikin', 'film', 'lebih', 'baik', 'kalo', 'penonton', 'dikit', 'mereka', 'rugi', 'pasti', 'dari', 'situ', 'mikir', 'apa', 'yang', 'salah', 'pls', 'lah', 'alur', 'kayak', 'gini', 'tuh', 'bagus', 'buat', 'bangun', 'nasionalisme']	tonton gaes bentuk mendukung negara support bikin film penonton dikit rugi situ mikir salah alur bagus bangun nasionalisme anak anak jaman	tonton gaes bentuk dukung negara support bikin film tonton dikit rugi situ mikir salah alur bagus bangun nasionalisme anak anak jaman

utk anak anak jaman skrng tp ga bgt	utk anak anak jaman skrng tp ga bgt	untuk anak anak jaman sekarang tapi tidak banget	'yang', 'salah', 'pls', 'lah', 'alur', 'kayak', 'begini', 'tuh', 'bagus', 'buat', 'bangun', 'nasionalisme', 'untuk', 'anak', 'anak', 'jaman', 'sekarang', 'tapi', 'tidak', 'banget']
---	---	---	---

3.4. Pembobotan TF-IDF

Setelah melakukan preprocessing data dari hasil *stemming* kemudian digunakan untuk proses selanjutnya pada tahapan *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode TF-IDF memeberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen serta tingkat keunikannya kata di seluruh dokumen [4]. Hasil pembobotan menunjukkan bahwa kata seperti “tonton” dan “anak” dalam dokumen memiliki nilai TF tinggi, sedangkan nilai IDF lebih rendah. Dari nilai TF-IDF tersebut menghasilkan bobot nilai yang merepresentasi fitur yang dihasilkan dapat digunakan pada tahap klasifikasi sentimen. Tabel 3 menampilkan hasil dari perhitungan TF-IDF.

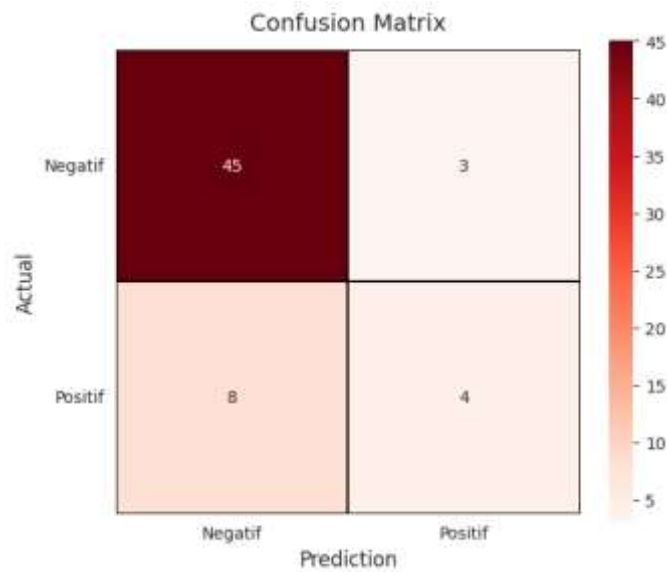
Tabel 4. Hasil Perhitungan TF-IDF

Kata	TF	IDF	TF-IDF
tonton	2	2.178976947	4.359753894
gaes	1	2.480006942	2.480006942
bentuk	1	2.480006942	2.480006942
dukung	1	2.480006942	2.480006942
negara	1	2.480006942	2.480006942
support	1	2.480006942	2.480006942
bikin	1	2.480006942	2.480006942
film	1	2.480006942	2.480006942
dikit	1	2.480006942	2.480006942
rugi	1	2.480006942	2.480006942
situ	1	2.480006942	2.480006942
mikir	1	2.480006942	2.480006942
salah	1	2.480006942	2.480006942
alur	1	2.480006942	2.480006942
bagus	1	2.480006942	2.480006942
bangun	1	2.480006942	2.480006942
nasionalisme	1	2.480006942	2.480006942
anak	2	2.178976947	4.359753894
jaman	1	2.480006942	2.480006942

TF-IDF sebagai metode *feature extraction* yang didasarkan pada efektifitas yang dilakukan pada penelitian [16] yang menerapkan TF-IDF dengan algoritma Support Vector Machine (SVM) pada analisis sentimen maskapai penerbangan menggunakan 2000 data hingga tingkat akurasi mencapai 84.37%. Pada penelitian tersebut metode Naïve Bayes dengan TF-IDF juga diterapkan dengan tingkat akurasi 78.21%.

3.5. Klasifikasi Naïve Bayes

Setelah melewati berbagai tahapan sebelumnya, Proses berikutnya melakukan klasifikasi dengan metode Naïve Bayes, data yang diperoleh pada media sosial X dengan jumlah 302 data, proses ini dilakukan dengan membagi data menjadi data latih dan data uji, lalu menggunakan fungsi *train_test_split* dengan parameter *test_size=0.2*, proses ini menghasilkan pembagian data menjadi 80% untuk data latih berjumlah 238 data dan 20% untuk data uji berjumlah 60 data. Hasil penerapan algoritma Naïve Bayes yang telah melewati berbagai tahapan seperti proses preprocessing dan pembobotan. Kemudian dari hasil klasifikasi melalui *Confusion Matrix*, model mengklasifikasi 45 data *True Negative* (data negatif yang diklasifikasi benar), serta 3 data *False Positive* (data negatif yang diprediksi sebagai positif), 4 data *True Positive* (data positif yang diprediksi benar positif), dan 8 data *False Negative* (data positif yang dalam diklasifikasi sebagai negatif). Hasil *Confusion Matrix* ditampilkan pada Gambar 4.



Gambar 3. Hasil Confusion Matrix

3.6. Evaluasi Model

Langkah terakhir dilakukan pada penelitian ini dengan melihat hasil kinerja model, dengan melakukan evaluasi menggunakan *Classification Report*. Nilai ini didapat berdasarkan hasil dari *Confusion Matrix*, yaitu menggunakan nilai *True Positive*, *True Negative*, *False Positive*, dan *False Negative*.

Classification Report:				
	precision	recall	f1-score	support
Negatif	0.85	0.94	0.89	48
Positif	0.57	0.33	0.42	12
accuracy			0.82	60
macro avg	0.71	0.64	0.66	60
weighted avg	0.79	0.82	0.80	60

Gambar 4. Hasil Classification Report

Hasil akurasi yang ditunjukkan dari klasifikasi model pada gambar 5 yaitu 0.82, dengan hasil akurasi dapat dihitung dengan menerapkan rumus (3) berikut ini.

$$Accuracy = \frac{4 + 45}{4 + 45 + 3 + 8} = \frac{49}{60} = 0.8167 \quad (3)$$

Kemudian menghitung nilai dari masing-masing kelas sentimen pada nilai *precision*, *recall* dan *f1-score* dengan menerapkan rumus (4), (5), dan (6) berikut ini.

$$Precision \text{ Positif} = \frac{4}{4 + 3} = \frac{4}{7} = 0.5714 \quad (4)$$

$$Precision \text{ Negatif} = \frac{45}{45 + 8} = \frac{45}{53} = 0.8490 \quad (4)$$

$$Recall \text{ Positif} = \frac{4}{4 + 8} = \frac{4}{12} = 0.3333 \quad (5)$$

$$Recall \text{ Negatif} = \frac{45}{45 + 3} = \frac{45}{48} = 0.9375 \quad (5)$$

$$F1\text{-Score Positif} = 2 \cdot \frac{0.57 \cdot 0.33}{0.57 + 0.33} = 2 \cdot 0.209 = 0.418 \quad (6)$$

$$F1\text{-Score Negatif} = 2 \cdot \frac{0.85 \cdot 0.94}{0.85 + 0.94} = 2 \cdot 0.4463 = 0.8927 \quad (6)$$

Berdasarkan hasil keseluruhan klasifikasi model serta dilakukan perhitungan secara manual, bahwa pada kelas sentimen negatif model memperoleh nilai *precision* sebesar 85%, *recall* sebesar 94%, *f1-score* sebesar 89%. Pada nilai *recall* menghasilkan akurasi tinggi, menunjukkan bahwa model memprediksi data sentimen dengan cukup baik. Sedangkan nilai pada kelas sentimen positif, model menghasilkan nilai *precision* sebesar 57%, *recall* sebesar 33%, dan *f1-score* sebesar 42%. Diketahui pada nilai *recall* yang rendah menandakan bahwa model kurang baik dalam mengidentifikasi data positif.

3.7. Wordcloud

Visualisasi *wordcloud* digunakan untuk melihat kata-kata di dalam kategori sentimen positif dan negatif [3]. Pada setiap gambar yang ditunjukkan, jika terdapat kata yang berukuran besar menandakan bahwa kata tersebut sering muncul dalam kategori tersebut.



Gambar 5. Wordcloud Positif

Hasil pada Gambar 6 yang menampilkan visualisasi *wordcloud* positif bahwa kata seperti bagus, keren, bikin, anak bangsa, karya, dan animasi. Terlihat bahwa kata-kata tersebut dominan muncul dalam kategori sentimen positif, ini menandakan adanya dukungan terhadap film animasi *Merah Putih One For All* dan apresiasi terhadap karya lokal menjadi faktor penting dalam sentimen positif.



Gambar 6. Wordcloud Negatif

Hasil visualisasi *wordcloud* juga tampilan pada Gambar 7 kategori sentimen negatif, terlihat kata-kata yang sering muncul pada sentimen negatif seperti jelek, poster, anak, animasi, grafik, dan konten. Kata tersebut muncul berdasarkan respon penonton terhadap kualitas film animasi *Merah Putih One For All*. Hasil menunjukkan bahwa komentar negatif ditujukan untuk mengkritik pada kualitas visual, desain poster, serta isi konten film yang dinilai kurang menarik. Seperti pada penelitian [4] visualisasi *wordcloud* berfokus pada hasil visualisasi dari keseluruhan kategori positif, negatif, dan netral. Sehingga pada penelitian ini menampilkan dan membedakan hasil *wordcloud* dari kategori sentimen positif dan sentimen negatif.

4. Kesimpulan

Hasil penelitian analisis sentimen terhadap film animasi "Merah Putih: One For All" menggunakan algoritma klasifikasi Naïve Bayes menunjukkan bahwa opini publik didominasi oleh sentimen negatif. Dari total 302 komentar yang dihimpun melalui platform X (dahulu Twitter), terdapat 240 komentar berkategori negatif dan 62 komentar berkategori positif. Penelitian ini menggunakan rasio pembagian data 80:20 untuk data latih dan data uji. Berdasarkan hasil pengujian, algoritma Naïve Bayes mampu memprediksi 53 komentar sebagai sentimen negatif dan 7 komentar sebagai sentimen positif pada data uji. Kinerja model menghasilkan nilai akurasi sebesar 82%, presisi 79%, *recall* 82%, dan *F1-score* 80%. Evaluasi tersebut menunjukkan bahwa algoritma Naïve Bayes memiliki performa yang baik dalam mengklasifikasikan sentimen, terutama dalam mengidentifikasi komentar negatif.

Penelitian selanjutnya disarankan untuk membandingkan kinerja algoritma Naïve Bayes dengan metode klasifikasi lain, seperti *Support Vector Machine* (SVM) atau *K-Nearest Neighbors* (KNN), guna memperoleh performa yang lebih optimal. Diperlukan pula penerapan teknik penyeimbangan data (*data balancing*) mengingat terdapat ketidakseimbangan distribusi antara kelas positif dan negatif. Selain itu, pengembangan penelitian dapat memperluas sumber dan volume data dari berbagai platform media sosial agar hasil analisis lebih representatif. Hasil penelitian ini diharapkan dapat menjadi bahan evaluasi bagi produser serta pembuat konten animasi nasional dalam meningkatkan kualitas produksi. Bagi praktisi pemasaran, temuan ini dapat digunakan untuk merancang strategi promosi yang lebih efektif guna mendukung pengembangan industri animasi di Indonesia.

5. Daftar Pustaka

- [1] F. S. Mufidah, S. Winarno, F. Alzami, E. D. Udayanti, and R. R. Sani, "Analisis Sentimen Masyarakat Terhadap Layanan ShopeeFood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier," *JOINS (Journal of Information System)*, vol. 7, no. 1, pp. 14–25, 2022, doi: 10.33633/joins.v7i1.5883.
- [2] Muhammad Rizal, M. Martanto, and U. Hayati, "Analisis Sentimen Pengguna Twitter Terkait Film One Piece Menggunakan Metode Naive Bayes," *Jurnal Sistem Informasi Kaputama (JSIK)*, vol. 8, no. 1, pp. 38–47, 2024, doi: 10.59697/jsik.v8i1.522.
- [3] Radendha Muhammad Arthansa, Dhea Intan Sagita, and Anggraini Puspita Sari, "Komparasi Analisis Sentimen Ulasan Film Avengers: Endgame Di Imdb Menggunakan Metode Naïve Bayes Dan Svm," *STORAGE – Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 3, no. 4, pp. 156–166, 2024.
- [4] Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes," *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 1, no. 4, pp. 411–423, 2022, doi: 10.55123/insologi.v1i4.770.
- [5] P. G. Aryanti and I. Santoso, "Analisis Sentimen Pada Twitter Terhadap Mobil Listrik Menggunakan Algoritma Naive Bayes," *IKRA-ITH Informatika : Jurnal Komputer dan Informatika*, vol. 7, no. 2, pp. 133–137, 2023, [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/2821>
- [6] A. R. Se, R. Amarullah, and R. Pribadi, "Analisis Sentimen Masyarakat Terhadap Film Animasi Jumbo Di Platform Tiktok Menggunakan Algoritma Naïve Bayes," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, pp. 181–187, 2025, doi: 10.23960/jitet.v13i3.6669.
- [7] A. Cahya Kamilla, N. Priyani, R. Priskila, and V. Handrianus Pranatawijaya, "Analisis Sentimen Film Agak Laen Dengan Kecerdasan Buatan: Text Mining Metode Naïve Bayes Classifier," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 2923–2928, 2024, doi: 10.36040/jati.v8i3.9587.
- [8] P. Al Muq̄sith Prasetyo and A. Hermawan, "Analisis sentimen twitter terhadap pemilihan presiden menggunakan algoritma naïve bayes," *INFOTECH : Jurnal Informatika & Teknologi*, vol. 4, no. 2, pp. 224–233, 2023, doi: 10.37373/infotech.v4i2.863.
- [9] A. K. Fauziyyah, "ANALISIS SENTIMEN PANDEMI COVID19 PADA STREAMING TWITTER DENGAN TEXT MINING PYTHON," *Jurnal Ilmiah SINUS*, vol. 18, no. 2, p. 31, Jul. 2020, doi: 10.30646/sinus.v18i2.491.
- [10] I. Verawati and B. S. Audit, "Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1411, Jul. 2022, doi: 10.30865/mib.v6i3.4132.
- [11] R. Wati, S. Ernawati, and H. Rachmi, "Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 13, no. 1, pp. 84–93, Apr. 2023, doi: 10.34010/jamika.v13i1.9424.

- [12] N. A. Maulana and D. Darwis, "Perbandingan Metode SVM dan Naïve Bayes untuk Analisis Sentimen pada Twitter tentang Obesitas di Kalangan Gen Z," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 3, pp. 655–666, Mar. 2025, doi: 10.52436/1.jpti.691.
- [13] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 542, Jan. 2022, doi: 10.30865/mib.v6i1.3530.
- [14] F. Zamzami, R. Hidayat, and R. Fathonah, "Penerapan Algoritma Naive Bayes Classifier Untuk Analisis Sentimen Komentar Twitter Proyek Pembangunan Ikn," *Faktor Exacta*, vol. 17, no. 1, pp. 47–57, 2024, doi: 10.30998/faktorexacta.v17i1.22265.
- [15] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *Jurnal Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [16] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, Feb. 2021, doi: 10.34148/teknika.v10i1.311.