

ANALISIS KOMPARATIF ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI SENTIMEN ULASAN WISATAWAN PADA DESTINASI GUNUNG MUTIS MELALUI PLATFORM TIKTOK

Yudiztira Syamsul Rizal, Erna Rosani Nubatonis, Heni

Teknik Informatika, STIKOM Uyelindo Kupang

Jl. Perintis Kemerdekaan 1, Kayu Putih, Kec.Oebobo, Kota Kupang, Nusa Tenggara Timur. 85228

yudiztirarizal020503@gmail.com

ABSTRAK

Perkembangan media sosial, khususnya TikTok, memungkinkan wisatawan menyampaikan opini dan pengalaman yang dapat dimanfaatkan sebagai sumber informasi dalam pengelolaan destinasi wisata seperti Gunung Mutis. Namun, data ulasan yang besar dan tidak terstruktur menyulitkan proses analisis secara manual. Oleh karena itu, penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) dalam klasifikasi sentimen ulasan wisatawan. Metode penelitian meliputi pengumpulan data melalui *web scraping*, *preprocessing data* (*case folding*, *cleansing*, *tokenizing*, *stopword removal*, dan *stemming*), pelabelan sentimen, serta ekstraksi fitur menggunakan TF-IDF. Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20, serta dilakukan penyeimbangan data menggunakan SMOTE. Evaluasi model dilakukan menggunakan metrik akurasi. Hasil penelitian menunjukkan bahwa *Naive Bayes* memperoleh akurasi sebesar 72,33%, sedangkan *Support Vector Machine* (SVM) mencapai akurasi sebesar 96,12%. Dengan demikian, SVM memiliki performa yang lebih baik dalam mengklasifikasikan sentimen ulasan wisatawan pada data TikTok.

Kata kunci : Analisis Sentimen, Machine Learning, Naive Bayes, Support Vector Machine, TF-IDF, TikTok.

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi, khususnya media sosial, telah mengubah cara masyarakat berinteraksi dan berbagi informasi secara *real-time*. TikTok menjadi salah satu platform yang menyediakan ruang bagi pengguna untuk membagikan opini dan ulasan, termasuk terkait destinasi wisata, yang dapat dimanfaatkan sebagai sumber data untuk memahami persepsi, kepuasan, dan pengalaman wisatawan.

Namun, data ulasan yang dihasilkan dari TikTok memiliki karakteristik berupa jumlah yang besar, dinamis, dan tidak terstruktur, sehingga menimbulkan tantangan dalam proses pengolahan dan analisis secara manual. Kondisi ini menyebabkan sulitnya memperoleh informasi yang akurat dan cepat sebagai dasar dalam pengambilan keputusan di bidang pariwisata.

Untuk mengatasi permasalahan tersebut, diperlukan metode analisis sentimen berbasis *machine learning* yang mampu mengklasifikasikan opini ke dalam kategori positif, negatif, dan netral secara otomatis. Algoritma yang umum digunakan adalah *Naive Bayes* dan *Support Vector Machine* (SVM), yang masing-masing memiliki kelebihan dan keterbatasan [1].

Naive Bayes bersifat *probabilistik* dengan proses yang sederhana dan efisien, sedangkan SVM menggunakan konsep margin maksimum yang umumnya menghasilkan akurasi lebih tinggi pada data berdimensi tinggi [2].

Berdasarkan hal tersebut, penelitian ini bertujuan untuk melakukan analisis komparatif antara algoritma

Naive Bayes dan *Support Vector Machine* (SVM) dalam klasifikasi sentimen ulasan wisatawan pada destinasi Gunung Mutis melalui platform TikTok.

Sebagai rencana penyelesaian masalah, penelitian ini dilakukan melalui tahapan pengumpulan data menggunakan *web scraping*, *preprocessing data*, pelabelan sentimen, ekstraksi fitur menggunakan TF-IDF, serta penerapan algoritma *Naive Bayes* dan SVM. Hasil dari kedua metode kemudian dibandingkan untuk menentukan model terbaik dalam mendukung pengelolaan destinasi wisata berbasis data. Penelitian ini juga diharapkan dapat berkontribusi dalam pengembangan teknologi informasi di bidang pariwisata serta menjadi referensi untuk pengembangan model analisis sentimen selanjutnya [3].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian oleh Supian *et al.* dalam *Jurnal Ilmiah Informatika (JIF)* menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memiliki akurasi lebih tinggi dibandingkan *Naive Bayes* dalam analisis sentimen terhadap isu pembangunan Ibu Kota Nusantara. Hasil ini menunjukkan bahwa SVM lebih unggul dalam menangani data berdimensi tinggi dengan distribusi fitur yang kompleks [1].

Selanjutnya, penelitian oleh Setyawan *et al.* dalam *Smart Comp: Jurnalnya Orang Pintar Komputer* membahas analisis sentimen ulasan wisata menggunakan *Naive Bayes* dan SVM. Hasil penelitian menunjukkan bahwa SVM memberikan hasil klasifikasi yang lebih stabil dan akurat dibandingkan

Naive Bayes, khususnya dalam konteks pariwisata berbasis ulasan digital [3].

Penelitian oleh Khasanah juga menunjukkan bahwa SVM memiliki keunggulan dalam hal akurasi, sementara *Naive Bayes* lebih unggul dalam kecepatan proses pelatihan dan pengujian. Hal ini menunjukkan bahwa kedua algoritma memiliki karakteristik yang berbeda dan dapat saling melengkapi dalam analisis sentimen [4].

2.2. Analisis Sentimen pada Media Sosial

Analisis sentimen merupakan bagian dari *Natural Language Processing* (NLP) yang bertujuan mengidentifikasi dan mengklasifikasikan opini atau emosi pengguna berdasarkan teks. Media sosial menjadi sumber data potensial karena pengguna aktif menyampaikan pendapat melalui komentar dan ulasan pada platform seperti TikTok, Instagram, dan Twitter (X) [5].

Dalam pariwisata digital, analisis sentimen berperan penting untuk memahami persepsi dan kepuasan wisatawan secara cepat dan objektif.

Namun, data media sosial bersifat tidak terstruktur serta memiliki tantangan seperti bahasa tidak baku, campuran bahasa, dan penggunaan emoji yang memengaruhi makna sentimen. Oleh karena itu, diperlukan pra-proses dan pemilihan algoritma yang tepat. Penelitian ini menggunakan *Naive Bayes* dan *Support Vector Machine* (SVM) untuk klasifikasi sentimen ulasan wisatawan pada destinasi Gunung Mutis di TikTok.

2.3. Klasifikasi Sentimen dan Algoritma Utama

Klasifikasi sentimen merupakan proses pengelompokan opini atau teks berdasarkan polaritas emosional (positif, negatif, atau netral) untuk memahami persepsi publik terhadap suatu topik. Dalam media sosial, metode ini digunakan untuk menganalisis opini pengguna melalui komentar, ulasan, maupun caption secara *real-time*.

Proses klasifikasi sentimen meliputi tahapan pra-pemrosesan (*case folding*, *tokenizing*, *stopword removal*, dan *stemming*), ekstraksi fitur menggunakan TF-IDF, klasifikasi dengan *Naive Bayes* dan *Support Vector Machine* (SVM), serta evaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*.

Beberapa penelitian menunjukkan bahwa SVM cenderung memiliki performa lebih baik dibanding *Naive Bayes*. Setiawan et al. menunjukkan SVM menghasilkan akurasi lebih tinggi pada analisis komentar TikTok [6].

Saputra et al. juga melaporkan akurasi SVM sebesar 88,85% dibanding *Naive Bayes* 72,64% [7]. Hasil serupa ditunjukkan oleh Agustin [8], serta Maulana & Darwis yang menemukan SVM lebih unggul dalam analisis sentimen media sosial [9].

2.4. Algoritma Naive Bayes

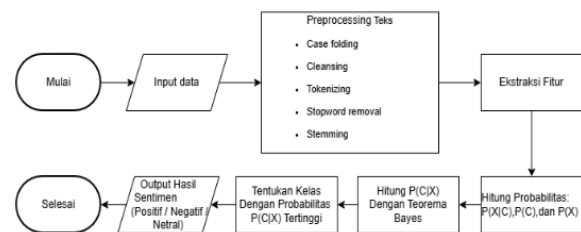
Naive Bayes merupakan algoritma berbasis probabilitas yang berlandaskan *Teorema Bayes* untuk

menghitung peluang suatu kelas berdasarkan bukti yang ada. Algoritma ini mengasumsikan bahwa setiap fitur bersifat independen (*independent feature assumption*) dan banyak digunakan dalam klasifikasi teks karena efisien dalam menghitung probabilitas kelas berdasarkan distribusi fitur pada data [10]. Rumus dasar Teorema Bayes adalah sebagai berikut:

$$P(C | X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Dalam implementasinya, *Naive Bayes* menghitung probabilitas setiap kelas berdasarkan frekuensi fitur pada data latih dan memilih kelas dengan probabilitas tertinggi sebagai hasil klasifikasi. Penelitian oleh Yudistira dan Isnain menunjukkan bahwa metode ini efektif dalam analisis sentimen karena mampu mengolah data teks secara cepat dengan akurasi yang baik [11].

Flowchart berikut menggambarkan alur kerja *Naive Bayes* mulai dari input data, perhitungan probabilitas, hingga penentuan kelas sentimen.



Gambar 1. Flowchart Algoritma *Naive Bayes*

2.5. Algoritma Support Vector Machine

Support Vector Machine (SVM) adalah algoritma *statistical learning* yang bekerja dengan mencari hyperplane optimal untuk memisahkan data ke dalam kelas yang berbeda dengan margin maksimum. Metode ini banyak digunakan dalam klasifikasi teks karena mampu menangani data berdimensi tinggi serta memberikan performa yang baik pada analisis sentimen [12]. Secara matematis, fungsi keputusan SVM dinyatakan sebagai:

$$f(x) = w \cdot x + b \quad (2)$$

Dengan w sebagai vektor bobot, x sebagai vektor input, dan b sebagai bias. Tujuan SVM adalah menentukan w dan b yang memaksimalkan margin antar kelas dengan formulasi optimasi:

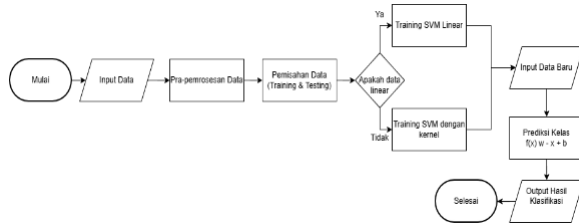
$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (3)$$

Dengan kendala:

$$y_i(w \cdot x_i + b) \geq 1, \forall i \quad (4)$$

Jika data tidak dapat dipisahkan secara linear, SVM menggunakan kernel seperti linear, polynomial, dan RBF untuk meningkatkan separabilitas data. Penelitian menunjukkan bahwa SVM memiliki performa yang baik dalam klasifikasi teks, terutama dengan kombinasi TF-IDF dan feature engineering, serta sering menghasilkan akurasi yang lebih tinggi dibanding metode lain pada analisis sentimen media sosial [13].

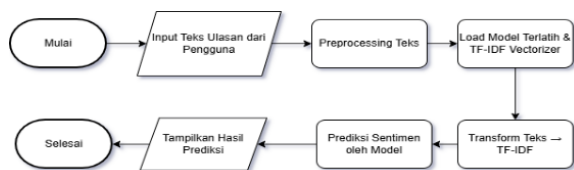
Berikut ditampilkan *flowchart* alur kerja SVM mulai dari proses *input* data hingga penentuan hasil klasifikasi.



Gambar 2. *Flowchart* Algoritma Support Vector Machine

3. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen. Alur penelitian ditunjukkan pada gambar berikut:



Gambar 3. *Flowchart* penelitian

3.1. Analisis Data

Analisis data dilakukan untuk memahami struktur dan kualitas data komentar TikTok tentang Gunung Mutis yang diperoleh melalui *scraping* menggunakan *Apify*. Data berupa *teks* tidak terstruktur sebanyak ±1.161 ulasan (2024–2025), disimpan dalam *Excel* dan diproses di *Google Colab* dengan *pandas*. Kualitas data ditingkatkan dengan menghapus duplikasi, membersihkan *teks*, dan mengevaluasi distribusi sentimen sebelum pemodelan.

3.2. Pengolahan Data

Pengolahan data dilakukan di *Google Colab* untuk menyiapkan data klasifikasi sentimen melalui tahapan preprocessing, pelabelan, pembagian dataset, ekstraksi fitur, pelatihan, evaluasi, dan perbandingan model. Preprocessing meliputi case folding, cleansing, tokenizing, stopword removal, dan stemming. Data diberi label secara manual dan lexicon-based, kemudian dibagi 80% training dan 20% testing. Fitur diekstraksi dengan TF-IDF, lalu model *Naive Bayes* dan *SVM* dilatih dan dievaluasi menggunakan accuracy, precision, recall, dan F1-score. Hasil perbandingan digunakan untuk menentukan model terbaik yang diimplementasikan pada aplikasi *Streamlit*.

3.3. Analisis Sistem

Analisis sistem dilakukan untuk memahami kebutuhan, alur kerja, dan komponen dalam pengembangan sistem klasifikasi sentimen ulasan wisatawan pada destinasi Gunung Mutis di platform TikTok. Sistem ini bertujuan untuk menganalisis

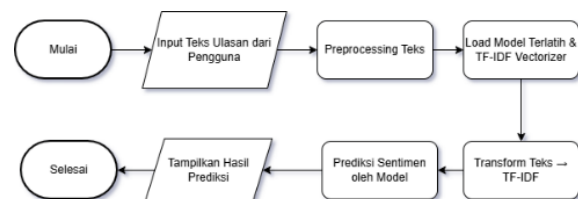
sentimen serta membandingkan kinerja algoritma *Naive Bayes* dan *Support Vector Machine* (SVM).

Data diperoleh melalui web *scraping* menggunakan *Apify* dan diolah menggunakan *Google Colab* berbasis *Python*. Proses sistem meliputi analisis kualitas data, *preprocessing*, pelabelan sentimen, pembagian dataset (80% data latih dan 20% data uji), serta ekstraksi fitur menggunakan TF-IDF. Selanjutnya dilakukan pelatihan dan evaluasi model *Naive Bayes* dan *SVM* untuk mengukur performa masing-masing algoritma. *Output* sistem berupa hasil klasifikasi sentimen, nilai evaluasi (*accuracy*, *precision*, *recall*, *F1-score*), perbandingan kinerja model, serta model terbaik yang diimplementasikan ke aplikasi web berbasis *Streamlit* melalui *Visual Studio Code*.

3.4. Perancangan Sistem

a. Perancangan Alur Sistem

Alur sistem digambarkan dalam *flowchart* yang mencakup seluruh tahapan penelitian dari pengumpulan data, preprocessing, pelabelan, pembagian dataset, TF-IDF, pelatihan dan evaluasi *Naive Bayes* dan *SVM*, hingga implementasi *Streamlit*. Berikut alur sistem aplikasi:



Gambar 4. *Flowchart* alur sistem

b. Perancangan Arsitektur Sistem

Arsitektur sistem meliputi TikTok sebagai sumber data, *Apify* untuk pengambilan data, *Google Colab* untuk pemodelan, dan *Visual Studio Code* untuk pengembangan aplikasi. Model terbaik diimplementasikan ke *Streamlit* untuk menampilkan hasil analisis sentimen.

c. Perancangan Perangkat Lunak dan Library

Sistem dibangun menggunakan *Python* dengan pustaka *pandas*, *scikit-learn*, serta *library preprocessing teks*. Antarmuka menggunakan *Streamlit*, sedangkan *pickle*, *pathlib*, dan *time* digunakan untuk penyimpanan model, manajemen file, dan pengaturan proses aplikasi.

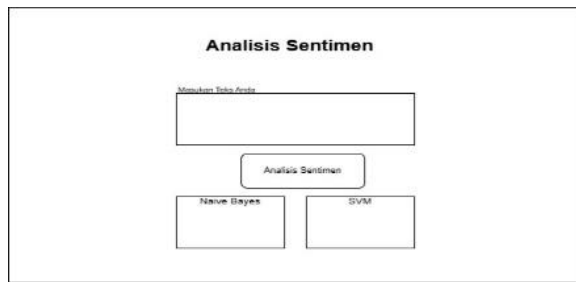
d. Perancangan Integrasi Model dan Implementasi Sistem

Model hasil pelatihan dan evaluasi di *Google Colab* disimpan kemudian diintegrasikan ke aplikasi web menggunakan *Visual Studio Code* dan *Streamlit*. Tahap ini bertujuan untuk mempermudah integrasi model, pengujian, serta proses *deployment* sistem.

e. Perancangan Antarmuka Pengguna

Antarmuka dirancang sederhana dan informatif agar mudah digunakan pengguna. Aplikasi *Streamlit* menampilkan hasil klasifikasi sentimen, ringkasan analisis, serta visualisasi sederhana

untuk membantu memahami distribusi sentimen. Berikut perancangan antarmuka pengguna:



Gambar 5. Perancangan antarmuka pengguna

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil komentar pengguna dari platform TikTok terkait wisata Gunung Mutis. Data yang diperoleh berupa teks komentar yang dikumpulkan dan disimpan sebagai dataset. Dataset awal masih berupa data mentah tanpa label sentimen, kemudian disimpan dalam format Microsoft Excel untuk memudahkan proses pengolahan.

#	text	label	replyCommentId	createTime	userId
1	semua kudu haranya hidup ket gini	4164		81 2025-03-12 11:26:00002	fitronha
2	alam tuh bagus bngt yah kkk ga drea	9568		10 2025-03-12 10:28:00002	rony2994
3	orang lain nantingga pakek kudu	11317		71 2025-03-12 13:36:00002	brangnabab
4	ga sempai ketemu orang lain di gunung	14298		39 2025-03-12 10:17:00002	blmrehabila
5	Kuda yg ada tanduk nya nantingga ket	1867		35 2025-03-12 10:58:00002	lgpncs3
6	dia lagi baik ke istana, aku tau soalnya aku ...	1991		9 2025-03-12 08:21:00002	kepermenbang
7	Indonesia bagian premium	28		0 2025-03-12 11:41:00002	elena_alred
8	Kuda yang putih lagi hant	2535		71 2025-03-12 13:36:00002	devianpaul006
9	Yaampun mereka lebih subur dari alam lain...	5367		10 2025-03-12 11:52:00002	demarungpaly
10	semua kudu haranya hidup ket gini	1919		4 2025-03-12 11:30:00002	rtvns
11	orang lain nantingga pakek kudu	75		0 2025-03-12 11:14:00002	msuaw33
12	Indonesia bagian premium	2756		70 2025-03-12 11:01:00002	brangnabab
13	Kuda yang putih lagi hant	4164		100 2025-03-12 14:26:00002	blmrehabila
14	Kuda yang putih lagi hant	1435		5 2025-03-12 14:33:00002	4ewens
15	Kuda yang putih lagi hant	925		0 2025-03-12 08:33:00002	adapito
16	Kuda yang putih lagi hant	416		4 2025-03-12 08:36:00002	viagapash
17	Kuda yang putih lagi hant	20979		40 2025-03-12 12:00:00002	a_m273
18	Kuda yang putih lagi hant	34		5 2025-03-12 11:50:00002	emyl1
19	Kuda yang putih lagi hant	424		0 2025-03-12 12:41:00002	rtvns
20	Kuda yang putih lagi hant	62		0 2025-03-12 12:54:00002	brangnabab
21	Kuda yang putih lagi hant	21		0 2025-03-12 12:20:00002	blmrehabila
22	Kuda yang putih lagi hant	622		3 2025-03-12 10:00:00002	setam_zavva
23	Kuda yang putih lagi hant	11		0 2025-03-12 13:36:00002	rtvns
24	Kuda yang putih lagi hant	31		0 2025-03-12 14:26:00002	devianpaul006
25	Kuda yang putih lagi hant	23		3 2025-03-12 12:35:00002	msuaw33

Gambar 6. menunjukkan contoh dataset komentar dalam format excel.

4.2. Preprocessing Data

Preprocessing data merupakan tahap penting untuk membersihkan dan menyiapkan teks sebelum klasifikasi Naive Bayes dan SVM agar lebih terstruktur dan optimal. Tahapan yang dilakukan meliputi case folding, cleaning, tokenizing, stopwords removal, dan stemming.

	text	case_folding	cleaning	tokenizing	stopword	stemming
0	semua kudu haranya hidup ket gini	semua kudu haranya hidup ket gini	semua kudu haranya hidup ket gini	[semua, kudu, haranya, hidup, ket, gini]	[kudu, hidup, ket, gini]	kudu hidup ket gini
1	alam tuh bagus bngt yah kkk ga drea	alam tuh bagus bngt yah kkk ga drea	alam tuh bagus bngt yah kkk ga drea	[alam, tuh, bagus, bngt, yah, kkk, ga, drea]	[alam, tuh, bagus, bngt, yah, kkk, ga, drea]	alam tuh bagus bngt yah kkk ga drea
2	orang lain nantingga pakek kudu	orang lain nantingga pakek kudu	orang lain nantingga pakek kudu	[orang, lain, nantingga, pakek, kudu]	[orang, lain, nantingga, pakek, kudu]	orang lain nantingga pakek kudu
3	ga sempai ketemu orang lain di gunung	ga sempai ketemu orang lain di gunung	ga sempai ketemu orang lain di gunung	[ga, sempai, ketemu, orang, lain, di, gunung]	[ga, sempai, ketemu, orang, lain, di, gunung]	ga sempai ketemu orang lain di gunung
4	Kuda yg ada tanduk nya nantingga ket	Kuda yg ada tanduk nya nantingga ket	Kuda yg ada tanduk nya nantingga ket	[kuda, yg, ada, tanduk, nya, nantingga, ket]	[kuda, yg, ada, tanduk, nya, nantingga, ket]	kuda yg ada tanduk nya nantingga ket
5	dia lagi baik ke istana, aku tau soalnya aku ...	dia lagi baik ke istana, aku tau soalnya aku ...	dia lagi baik ke istana, aku tau soalnya aku ...	[dia, lagi, baik, ke, istana, aku, tau, soalnya, aku]	[dia, lagi, baik, ke, istana, aku, tau, soalnya, aku]	dia lagi baik ke istana, aku tau soalnya aku
6	Indonesia bagian premium	Indonesia bagian premium	Indonesia bagian premium	[Indonesia, bagian, premium]	[Indonesia, bagian, premium]	Indonesia bagian premium
7	Kuda yang putih lagi hant	Kuda yang putih lagi hant	Kuda yang putih lagi hant	[kuda, yang, putih, lagi, hant]	[kuda, yang, putih, lagi, hant]	kuda yang putih lagi hant
8	Yaampun mereka lebih subur dari alam lain...	Yaampun mereka lebih subur dari alam lain...	Yaampun mereka lebih subur dari alam lain...	[Yaampun, mereka, lebih, subur, dari, alam, lain]	[Yaampun, mereka, lebih, subur, dari, alam, lain]	Yaampun mereka lebih subur dari alam lain
9	semua kudu haranya hidup ket gini	semua kudu haranya hidup ket gini	semua kudu haranya hidup ket gini	[semua, kudu, haranya, hidup, ket, gini]	[semua, kudu, haranya, hidup, ket, gini]	semua kudu haranya hidup ket gini

Gambar 1. Tahap preprocessing Data

4.3. Pelabelan Data

Pelabelan data adalah proses pengelompokan teks hasil preprocessing ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Proses ini menggunakan pendekatan lexicon-based berdasarkan

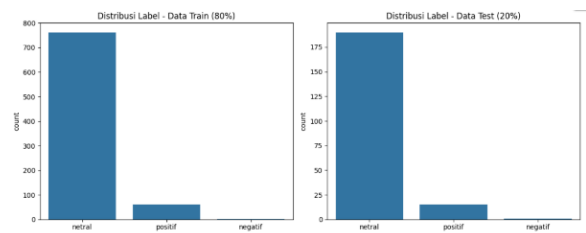
nilai *lex_score*, dengan ketentuan skor > 0 positif, < 0 negatif, dan = 0 netral.

	text	stemming	lex_score	label
661	kok gw liri ya sama kuda nya udah bisa rasain d...	gw liri ya kuda nya udah rasain d...	2	positif
778	semoga Allah melindungi Indonesia kita yang indah	moga allah lindung indonesia indah	1	positif
134	lali gini jd sedih teringat kuda2 yg dijadin ...	lali gini jd sedih kuda yg dijadin lalan kerja...	-2	negatif
24	jadi sedih di sini kuda pake narik delman ...	sedih kuda pake narik delman peotir huan an na...	-1	negatif
104	pangeran gantengnya mana bang?	pangeran ganteng bang	0	netral
100	keren bgtn kaya di negeri dongeng	keren bgtn kaya negeri dongeng	0	netral

Gambar 2. Pelabelan Data

4.4. Pembagian Data (Train-Test Split)

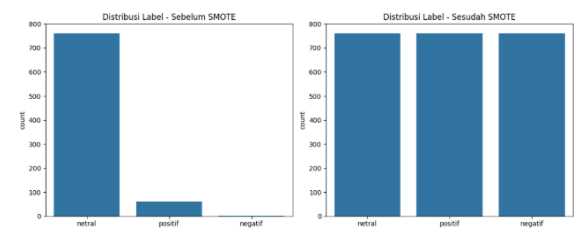
Pada tahap ini, dataset dibagi menjadi data training dan testing menggunakan train-test split dengan rasio 80% dan 20%. Parameter stratify digunakan untuk menjaga keseimbangan kelas. Berikut perbandingan jumlah data training dan testing:



Gambar 3. Tabel perbandingan jumlah data training dan test

4.5. Synthetic Minority Over-sampling Technique (SMOTE)

Pada tahap ini dilakukan penyeimbangan data menggunakan SMOTE untuk mengatasi ketidakseimbangan kelas, di mana kelas netral dominan dan kelas negatif paling sedikit. SMOTE hanya diterapkan pada data training setelah TF-IDF, sedangkan data testing tetap asli agar evaluasi tetap objektif. Berikut perbandingan distribusi data sebelum dan sesudah SMOTE:

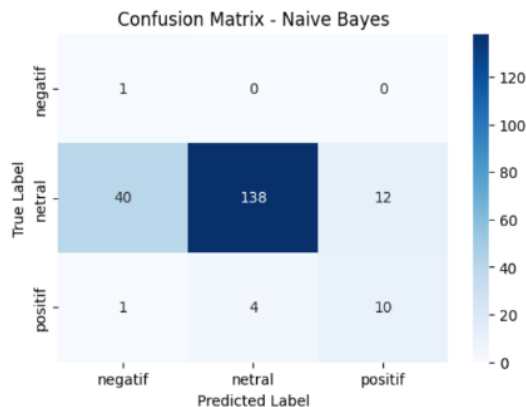


Gambar 4. Perbandingan distribusi data sebelum dan sesudah SMOTE

4.6. Implementasi Algoritma Naive Bayes

Pada tahap ini dilakukan implementasi algoritma Naive Bayes untuk klasifikasi sentimen ke dalam kelas negatif, netral, dan positif. Data terlebih dahulu diubah menjadi bentuk numerik menggunakan TF-IDF, kemudian dibagi menjadi 80% data training dan 20% data testing dengan stratify untuk menjaga keseimbangan kelas. Model yang digunakan adalah Complement Naive Bayes yang sesuai untuk data tidak seimbang. Setelah pelatihan pada data training, model digunakan untuk memprediksi data testing. Berikut classification report dan confusion matrix:

Classification Report:				
	precision	recall	f1-score	support
negatif	0.02	1.00	0.05	1
netral	0.97	0.73	0.83	190
positif	0.45	0.67	0.54	15
accuracy			0.72	206
macro avg	0.48	0.80	0.47	206
weighted avg	0.93	0.72	0.81	206

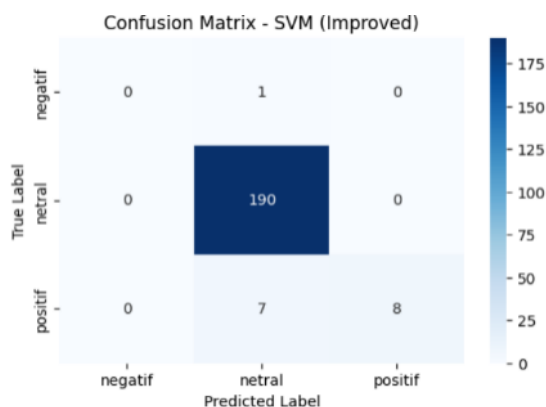
Gambar 5. Hasil Klasifikasi *Naive Bayes*Gambar 6. *Confusion Matrix Naive Bayes*

4.7. Implementasi Algoritma *Support Vector Machine*

Pada tahap ini digunakan SVM untuk klasifikasi sentimen negatif, netral, dan positif. Data diubah menjadi numerik dengan TF-IDF, lalu dibagi 80% *training* dan 20% *testing*. Ketidakseimbangan data pada training diatasi dengan SMOTE, sedangkan data testing tetap asli. Model SVM *kernel linear* dengan *class_weight* dilatih dan diuji menggunakan data *testing*, kemudian dievaluasi dengan *classification report* dan *confusion matrix*.

Classification Report:				
	precision	recall	f1-score	support
negatif	0.00	0.00	0.00	1
netral	0.96	1.00	0.98	190
positif	1.00	0.53	0.70	15
accuracy			0.96	206
macro avg	0.65	0.51	0.56	206
weighted avg	0.96	0.96	0.95	206

Gambar 7. Hasil Klasifikasi SVM

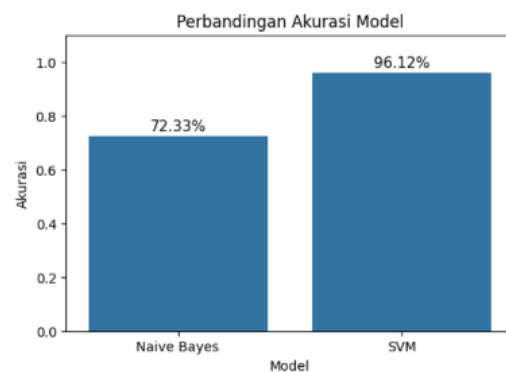
Gambar 8. *Confusion Matrix SVM*

4.8. Perbandingan Hasil *Naive Bayes* dan SVM

Pada tahap ini dilakukan perbandingan performa *Naive Bayes* dan *Support Vector Machine* (SVM) untuk mengetahui model terbaik berdasarkan data testing. Evaluasi menggunakan metrik akurasi yang dihitung dengan *accuracy_score* dari *scikit-learn*. Berikut hasil tabel perbandingan akurasi dan grafik perbandingan akurasi:

```

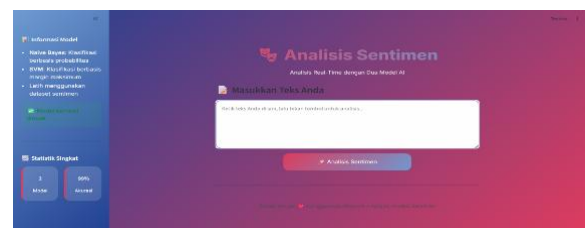
=== TABEL PERBANDINGAN AKURASI ===
Model    Akurasi
0 Naive Bayes  0.723301
1 SVM        0.961165
  
```

Gambar 9. Tabel Perbandingan Akurasi *Naive Bayes* dan SVMGambar 10. Grafik Perbandingan Akurasi *Naive Bayes* dan SVM

Hasil pengujian menunjukkan akurasi *Naive Bayes* sebesar 72,33% dan SVM sebesar 96,12%. Perbedaan ini dipengaruhi karakteristik algoritma dan distribusi data, di mana *Naive Bayes* lebih stabil pada data tidak seimbang, sedangkan SVM lebih efektif dalam pemisahan kelas sehingga menghasilkan akurasi lebih tinggi. Secara keseluruhan, SVM memberikan performa terbaik.

4.9. Implementasi Aplikasi Web *Streamlit*

Pada tahap ini sistem diimplementasikan sebagai aplikasi web menggunakan *Streamlit* untuk memudahkan analisis sentimen berdasarkan model yang telah dilatih. Aplikasi menyediakan fitur *input teks* dan menampilkan hasil klasifikasi (positif, negatif, atau netral) melalui proses *preprocessing*, TF-IDF, dan prediksi model.

Gambar 11. Tampilan Antarmuka Aplikasi *Streamlit*

4.10. Kelebihan Sistem

Sistem ini menggunakan *Naive Bayes* dan SVM untuk membandingkan performa dan menentukan model terbaik. Ketidakseimbangan data ditangani dengan SMOTE agar kelas minoritas dapat dipelajari dengan lebih baik. Data diproses melalui *preprocessing* (*case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*) agar lebih bersih dan konsisten. Sistem diimplementasikan menggunakan *Streamlit* agar mudah digunakan, serta dievaluasi dengan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix* untuk menilai performa klasifikasi secara menyeluruh.

4.11. Kekurangan Sistem

Sistem memiliki beberapa keterbatasan, yaitu masih sulit mengenali kelas minoritas meskipun sudah menggunakan SMOTE, terutama pada kelas negatif, serta dipengaruhi oleh dataset yang relatif kecil dan tidak seimbang. Selain itu, model memerlukan proses retraining jika terdapat pola data baru untuk menjaga akurasi. Analisis juga hanya terbatas pada teks komentar tanpa mempertimbangkan konten multimedia TikTok, dan performa sangat bergantung pada representasi TF-IDF sehingga perubahan pola kata dapat menurunkan akurasi.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, sistem berhasil mengklasifikasikan sentimen komentar TikTok pada destinasi Gunung Mutis ke dalam tiga kategori sehingga tujuan penelitian tercapai. SVM memperoleh akurasi lebih tinggi yaitu 96,12% dibanding *Naive Bayes* 72,33%, karena lebih efektif dalam menangani data kompleks dan didukung SMOTE untuk mengatasi ketidakseimbangan data, sedangkan *Naive Bayes* lebih sederhana namun kurang optimal dalam menangkap pola teks. Data didominasi sentimen netral (762), positif (60), dan negatif (2) sehingga diperlukan SMOTE agar model lebih seimbang. Implementasi *Streamlit* mempermudah analisis secara interaktif, dan hasil penelitian dapat dimanfaatkan untuk evaluasi kepuasan serta pengembangan dan promosi wisata Gunung Mutis..

Berdasarkan hasil penelitian, disarankan penambahan dan variasi dataset terutama pada kelas minoritas serta perluasan sumber data dari berbagai platform media sosial. Selain itu, dapat digunakan metode yang lebih kompleks seperti *deep learning* (LSTM atau Transformer) atau *word embedding* (Word2Vec, FastText, BERT) untuk meningkatkan akurasi dan pemahaman konteks. Sistem juga dapat dikembangkan dengan fitur *retraining* otomatis serta diperluas untuk analisis data multimedia. Hasil penelitian diharapkan dapat dimanfaatkan *stakeholder* untuk evaluasi dan pengambilan keputusan dalam pengembangan dan promosi wisata Gunung Mutis.

DAFTAR PUSTAKA

- [1] A. Supian, B. Tri Revaldo, N. Marhadi, L. Efrizoni, and R. Rahmaddeni, "Perbandingan Kinerja Naive Bayes Dan Svm Pada Analisis Sentimen Twitter Ibukota Nusantara," *J. Ilm. Inform.*, vol. 12, no. 01, pp. 15–21, 2024, doi: 10.33884/jif.v12i01.8721.
- [2] W. Ningsih, B. Alfiana, R. Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.
- [3] Y. Anggit Setyawan, F. Ely Nastiti, and A. Arum Sari, "Analisis Sentimen Ulasan Tempat Wisata Umbul Sigedang Di Google Maps Menggunakan Algoritma Naive Bayes," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 14, no. 2, 2025, doi: 10.30591/smartcomp.v14i2.7467.
- [4] Uswatun Khasanah, "Analisis sentimen terhadap tempat wisata menggunakan metode naive bayes classifier dan Support Vector Machine," pp. 1–112, 2023.
- [5] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naive bayes dan Support Vector Machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.
- [6] S. Liem, T. Setiawan, and M. R. Pribadi, "Perbandingan Algoritma SVM dan Naive Bayes Berbasis SMOTE dalam Analisis Sentimen Komentar Tiktok pada Produk Skincare," *Appl. Inf. Technol. Comput. Sci.*, vol. 3, no. 1, pp. 28–32, 2024, doi: 10.58466/aicoms.v3i1.1523.
- [7] J. Saputra, L. Maryani, Rahmaddeni, D. Wulandari, and W. Eka, "Analisis Performa Naive Bayes Dan Svm Terhadap Sentimen Teks Media Sosial Dengan Word2Vec Dan Smote," *J. INSTEK (Informatika Sains dan Teknol.*, vol. 10, no. 1, pp. 143–155, 2025, doi: 10.24252/instek.v10i1.54889.
- [8] R. I. Agustin, "Komparasi Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Twitter Korupsi Bansos Beras Masa Pandemi," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 912–918, 2024, doi: 10.23960/jitet.v12i2.4020.
- [9] N. A. Maulana and D. Darwis, "Perbandingan Metode SVM dan Naive Bayes untuk Analisis Sentimen pada Twitter tentang Obesitas di Kalangan Gen Z," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 3, pp. 655–666, 2025, doi: 10.52436/1.jpti.691.
- [10] Yusuf Ramadhan Nasution, Suhardi Suhardi, and Ilham Hafiz Satrio, "Penerapan Algoritma Klasifikasi Naive Bayes Untuk Analisis Sentimen Tentang Pemilu 2024," *Elkom J. Elektron. dan Komput.*, vol. 17, no. 2, pp. 495–

- 502, 2024, doi: 10.51903/elkom.v17i2.2053.
- [11] F. Yudistira and A. Rahman Isnain, “Analisis Sentimen Terhadap Seleksi CPNS Tahun 2024 Berbasis Media Sosial X Menggunakan Algoritma Naïve Bayes SENTIMENT ANALYSIS OF THE 2024 CPNS SELECTION BASED ON SOCIAL MEDIA X USING THE NAÏVE BAYES ALGORITHM,” *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 3, pp. 887–897, 2025, [Online]. Available: <https://doi.org/10.52436/1.jpti.731>
- [12] S. Ratnaswari, N. C. Wibowo, and D. S. Y. Kartika, “Analisis Sentimen Menggunakan Metode Lexicon-Based Dan *Support Vector Machine* Pada Presiden Dan Wakil Presiden Indonesia Periode 2024–2029,” *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 1, pp. 362–368, 2025, doi: 10.23960/jitet.v13i1.5604.
- [13] M. Ma’rufudin and A. Yudhistira, “Analisis Sentimen Petani Milenial Pada Media Sosial X Menggunakan Algoritma *Support Vector Machine* (SVM),” *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 3, pp. 845–857, 2025, doi: 10.52436/1.jpti.717.