

Analisis Sentimen Vtuber Hololive Indonesia Sebagai Tren Hiburan Digital Menggunakan Naïve Bayes Dan Support Vektor Machine

1st Frederick Rianto Putra

Informatika

Telkom University Purwokerto

Purwokerto, Indonesia

frederickriantop@student.telkomuniversity.ac.id

2nd Paradise

Informatika

Telkom University Purwokerto

Purwokerto, Indonesia

paradise@telkomuniversity.ac.id

3rd Yohani Setiya Rafika Nur

Informatika

Telkom University Purwokerto

Purwokerto, Indonesia

yohanin@telkomuniversity.ac.id

Abstrak--Virtual YouTuber (VTuber) telah menjadi industri hiburan digital dengan signifikansi budaya dan ekonomi yang kuat, salah satunya diwakili oleh Hololive Indonesia (HoloID). Namun, dibalik pertumbuhan pesat dan interaksi penggemar yang masif, sering kali muncul polarisasi opini publik yang tajam, sehingga pemahaman yang terukur mengenai sentimen audiens menjadi krusial bagi manajemen citra dan strategi konten. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap HoloID di media sosial (YouTube dan Twitter) serta mengevaluasi dan membandingkan performa algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam tugas klasifikasi tersebut. Data diperoleh melalui crawling dan scraping, diikuti pre-processing untuk pembersihan dan penyesuaian data. Proses labeling dilakukan semi-otomatis menggunakan SenticNet, mengklasifikasikan sentimen menjadi positif dan negatif. Sentimen netral dihilangkan dari dataset akhir untuk memfokuskan analisis pada dua kutub opini utama. Dataset dibagi 80% untuk data latih dan 20% untuk data uji, serta dievaluasi dengan 10-fold cross-validation. Hasil evaluasi menunjukkan akurasi model Naïve Bayes tanpa SMOTE adalah 0.8022, dan dengan SMOTE 0.7997. Sementara itu, akurasi model SVM tanpa SMOTE mencapai 0.8980, dan dengan SMOTE 0.8926. Rata-rata 10-fold cross-validation menunjukkan akurasi tertinggi pada SVM tanpa SMOTE (0.9040), menjadikannya model terbaik dalam penelitian ini.

Kata kunci—Analisis Sentimen, Naïve Bayes, Support Vektor Machine

I. PENDAHULUAN

Fenomena Virtual YouTuber (VTuber) telah berkembang menjadi sebuah industri hiburan digital dengan signifikansi ekonomi dan budaya yang kuat. Kompleksitas fenomena ini salah satunya tercermin dari perdebatan publik yang intens mengenai harga tiket konser salah satu VTuber di Thailand, di mana diskusi di media sosial tidak hanya berfokus pada kelayakan harga, tetapi juga persepsi nilai VTuber sebagai hiburan yang disamakan dengan "menonton kartun bergerak". Peristiwa ini merepresentasikan dua tren kunci: komodifikasi pesat industri VTuber dan munculnya polarisasi opini publik yang tajam. Mengingat status Hololive Indonesia (HoloID) sebagai agensi VTuber terkemuka dengan basis penggemar masif di Indonesia, dinamika serupa berpotensi memengaruhi citra merek dan strategi monetisasi.

Oleh karena itu, pemahaman terukur mengenai sentimen publik menjadi krusial.

Seiring perkembangan teknologi seperti *motion capture*, VTuber merevolusi interaksi kreator-penggemar. Meskipun popularitasnya terus meningkat, belum ada pemahaman komprehensif mengenai bagaimana sentimen publik terbentuk di media sosial. Analisis sentimen menjadi penting untuk mengisi kekosongan tersebut, dengan mengevaluasi efektivitas algoritma *machine learning*. , VTuber dipilih sebagai fokus karena statusnya sebagai fenomena inovatif yang menjembatani teknologi animasi dengan interaksi manusia yang otentik, menciptakan kedekatan parasosial yang unik dan kuat. Fenomena ini bukan tren sesaat, melainkan pergeseran signifikan dalam lanskap media digital yang kaya data untuk dianalisis[1][2], [3].

Bermula dari pionir seperti Kizuna Ai, yang popularitasnya mendunia dan bahkan menjadi duta pariwisata Jepang serta hadir di acara di Indonesia, tren VTuber menyebar luas. Keberhasilan ini menjadi katalisator bagi agensi seperti Hololive, yang menggabungkan konsep idola dengan teknologi digital untuk menciptakan interaksi emosional dengan audiens. Hololive, yang dikembangkan oleh Cover Corporation, memperluas jangkauannya secara global, termasuk Hololive Indonesia (HoloID) sebagai cabang pertama di Asia Tenggara. Dengan memanfaatkan budaya pop Jepang dan tingginya penggunaan media sosial, HoloID berhasil membangun basis penggemar yang kuat, seperti yang ditunjukkan oleh pertumbuhan *subscriber* kanalnya[4].

Analisis sentimen pada platform seperti Twitter (X) dan YouTube, yang merupakan platform interaksi utama HoloID, dapat memberikan wawasan tentang persepsi publik. Observasi menunjukkan adanya perdebatan signifikan mengenai aspek komersial VTuber, yang menimbulkan pertanyaan tentang bagaimana sentimen publik terbentuk secara keseluruhan. Penelitian sebelumnya[5] oleh telah mengkaji sentimen terhadap VTuber menggunakan Support Vector Machine (SVM) dan SMOTE dengan akurasi 88,18%. Penelitian ini bertujuan mengisi kesenjangan pemahaman dengan menganalisis sentimen terhadap HoloID menggunakan perbandingan metode Naïve Bayes, yang efisien untuk data teks, dan SVM, yang andal dalam pemisahan kelas sentimen.

Berdasarkan latar belakang tersebut, penelitian ini merumuskan pertanyaan: Berapa akurasi algoritma Naïve

Bayes dan SVM dalam mengklasifikasikan sentimen publik terhadap Hololive Indonesia, dan bagaimana hasil perbandingan keduanya?. Tujuan dari penelitian ini adalah untuk menganalisis sentimen publik terhadap HoloID di YouTube dan X (Twitter) serta mengevaluasi dan membandingkan akurasi kedua algoritma tersebut dalam mengklasifikasikan sentimen.

II. KAJIAN TEORI

A. Sentimen Analisis

Analisis sentimen adalah disiplin komputasional di persimpangan Natural Language Processing (NLP) dan Machine Learning (ML) yang dirancang untuk mengurai, mengklasifikasi, dan menguantifikasi opini dalam data teksual. Teknik ini bekerja dengan mengenali isyarat linguistik untuk secara otomatis mengevaluasi polaritas pandangan penulis (positif, negatif, atau netral). Pendekatan ini sangat relevan untuk mengolah data tak terstruktur dari sumber digital seperti media sosial untuk memetakan persepsi publik[5]–[8].

B. Vtuber (Virtual Youtuber)

Virtual YouTuber (VTuber) adalah fenomena sosiokultural di mana seorang kreator konten digital menggunakan avatar virtual, biasanya karakter 2D atau 3D bergaya animasi. Avatar ini digerakkan secara real-time melalui teknologi motion capture, menciptakan entitas hibrida yang unik antara persona manusia otentik dan identitas digital fiktional. Platform utama yang digunakan untuk siaran langsung dan interaksi adalah YouTube dan Twitch[1], [9], [10].

C. Data Mining

Data mining adalah proses komputasional untuk mengekstraksi informasi dan pola tersembunyi dari kumpulan data berskala besar. Tujuannya adalah menemukan korelasi, anomali, atau tren yang valid dan bermanfaat. Tugas dalam data mining diklasifikasikan menjadi dua kategori utama: deskriptif (merangkum karakteristik data) dan prediktif (membangun model untuk memperkirakan nilai di masa mendatang). Proses data mining meliputi beberapa tahapan: Data Cleaning (menangani noise dan data yang hilang), Data Integration (konsolidasi data dari berbagai sumber), Data Selection (pemilihan data yang relevan), Data Transformation (mengubah data ke format yang sesuai), Data Mining (penerapan algoritma untuk mengekstraksi pola), Pattern Evaluation (validasi pola yang ditemukan), dan Knowledge Presentation (penyajian wawasan kepada pengguna akhir)[11].

D. Algoritma SVM

Support Vector Machine (SVM) adalah algoritma supervised learning yang dikembangkan oleh Vladimir Vapnik, yang bekerja dengan menemukan hyperplane optimal untuk memisahkan titik data ke dalam kelas yang berbeda. Prinsip fundamentalnya adalah maksimisasi margin, di mana SVM memilih hyperplane dengan jarak terjauh ke titik data terdekat dari setiap kelas (support vectors). SVM mampu melakukan klasifikasi non-linear dengan memetakan data ke ruang dimensi yang lebih tinggi dan hanya dapat digunakan untuk klasifikasi biner (dua kelas)[5], [7], [11], [12]. Persamaan matematis dari hyperplane dapat dituliskan sebagai berikut:

1. Rumus untuk Hyperplane

$$w \cdot x + b = 0$$

Keterangan:

w: Vektor berat yang menentukan arah hyperplane

x: Vektor data

b: Bias term, yang menentukan posisi hyperplane pada sumbu y

2. Rumus untuk fungsi objektif

$$\min \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i$$

Keterangan:

w: vektor bobot

ξ_i : variabel slack yang mengukur seberapa jauh sampel ke-i berada dari margin yang diharapkan.

C: parameter regularisasi yang mengontrol optimasi antara margin dan kesalahan klasifikasi

E. Algoritma Naïve Bayes

Naïve Bayes adalah sebuah metode klasifikasi probabilistik yang bekerja berdasarkan prinsip fundamental dari Teorema Bayes. Keunikan algoritma ini terletak pada asumsi “naïf” bahwa setiap fitur dalam data bersifat independen satu sama lain dalam menentukan hasil klasifikasi. Meskipun asumsi ini merupakan penyederhanaan, Naïve Bayes dikenal sangat efisien, cepat, dan sering menunjukkan kinerja yang sangat baik untuk tugas klasifikasi teks, menjadikannya model pembanding yang kuat dalam penelitian ini[12]–[16]. Rumus yang digunakan dalam algoritma ini adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Dimana :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas Hipotesis H berdasarkan kondisi X (posterior probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

III. METODE

Memberikan Metodologi penelitian ini dirancang untuk menganalisis sentimen publik terhadap VTuber Hololive Indonesia (HoloID) menggunakan algoritma Naïve Bayes dan SVM. Alur penelitian mencakup pengumpulan data, pre-processing, pelabelan, perancangan model, dan evaluasi.

1. Pengumpulan Data

Langkah awal adalah observasi sistematis terhadap konten di YouTube dan utas diskusi di Twitter (X) untuk mengidentifikasi sumber data yang relevan dengan topik Hololive Indonesia. Data sentimen publik kemudian dikumpulkan dari kedua platform tersebut melalui API Twitter dan scraping komentar YouTube, dengan total 16.134 komentar. Pengumpulan data juga didukung oleh studi dokumentasi berupa pengumpulan teks komentar dan tweet, serta tangkapan layar untuk memperkuat konteks analisis.

2. Pre-Processing Data

Data mentah yang telah dikumpulkan kemudian melalui serangkaian tahapan pre-processing untuk membersihkan noise dan menstrukturkan data. Tahapannya meliputi:

- a. **Case Folding:** Mengubah semua teks menjadi huruf kecil.
- b. **Data Cleaning:** Menghapus URL, simbol, angka, dan tanda baca yang tidak relevan.
- c. **Translation:** Menerjemahkan semua teks ke dalam satu bahasa yang konsisten untuk analisis.
- d. **Normalization Emoticon:** Menyamakan bentuk ekspresi seperti emotikon ke dalam representasi teks yang seragam.
- e. **Tokenizing:** Memecah teks menjadi unit-unit kata atau token.
- f. **Stopword Removal:** Menghilangkan kata-kata umum yang tidak memiliki makna signifikan (misalnya, kata penghubung).
- g. **Stemming:** Mengubah kata ke bentuk dasarnya untuk mengurangi variasi bentuk kata.

3. Labeling Data

Setelah pre-processing, data diberi label sentimen (positif dan negatif) secara semi-otomatis menggunakan SenticNet, sebuah kamus leksikon berbasis pengetahuan afektif. Untuk memastikan validitas pelabelan, sampel hasil diverifikasi oleh seorang ahli di bidang bahasa, yaitu Bapak Dhanu Priyo Widodo, S.Kom., M.Kom.

4. Perancangan Model

Dataset yang telah bersih dan berlabel dibagi menjadi 80% data latih dan 20% data uji. Model klasifikasi dibangun menggunakan dua algoritma: Naïve Bayes dan Support Vector Machine (SVM). Model Naïve Bayes dibangun dengan pendekatan probabilistik, sementara model SVM dibangun untuk menemukan hyperplane optimal yang memisahkan kelas sentimen positif dan negatif.

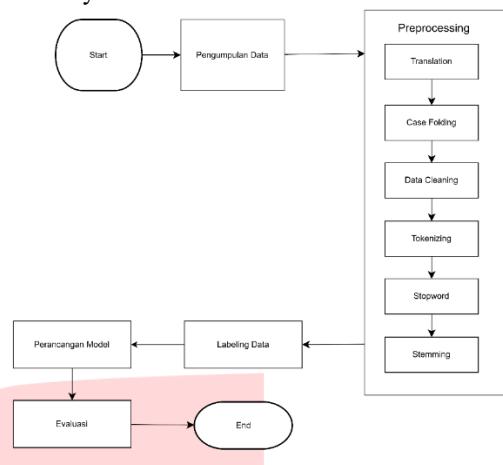
5. Evaluasi

Kinerja kedua model dievaluasi menggunakan metrik yang dihitung dari confusion matrix, yaitu akurasi, presisi, recall, dan F1-score. Selain itu, dilakukan validasi silang (0-fold cross-validation) untuk memastikan keakuratan dan stabilitas model.

IV. HASIL DAN PEMBAHASAN

Bagian ini berisi paparan objektif peneliti terhadap hasil-hasil penelitian berupa penjelasan dan analisis terhadap penemuan-penemuan penelitian, penjelasan serta penafsiran dari data dan hubungan yang diperoleh, serta pembuatan

generalisasi dari penemuan. Apabila terdapat hipotesis, maka pada bagian ini juga menjelaskan proses pengujian hipotesis beserta hasilnya.



GAMBAR 1
ALUR PEMODELAN

A Skenario Percobaan

Skenario percobaan dalam penelitian ini dirancang untuk menguji efektivitas algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan sentimen positif dan negatif terhadap VTuber Hololive Indonesia. Data sentimen diperoleh dari platform media sosial Twitter (X) dan YouTube. Proses dimulai dengan pengumpulan data, diikuti tahapan pre-processing dan labeling. Dataset yang telah bersih dan berlabel kemudian dibagi menjadi data latih dan data uji dengan beberapa rasio 80:20. Dan dilakukan juga Validasi model menggunakan 10-fold cross-validation untuk memastikan stabilitas dan generalisasi performa. Untuk mengatasi ketidakseimbangan kelas dalam dataset, diterapkan teknik Synthetic Minority Oversampling Technique (SMOTE). Metrik evaluasi yang digunakan untuk mengukur kinerja model meliputi akurasi, presisi, recall, dan F1-score, yang dihitung berdasarkan confusion matrix.

B Pengumpulan Data

Data dikumpulkan dari platform media sosial Twitter (X) dan YouTube berdasarkan topik-topik yang relevan dengan Hololive Indonesia. Kata kunci seperti nama-nama talent, hashtag acara (#holoIDcaf3, #GebyarholoID), serta aktivitas livestream digunakan untuk melakukan scraping dari Januari 2023 hingga Maret 2025. Data dikumpulkan menggunakan Google Colab dan library Python seperti snscreape untuk Twitter dan API YouTube untuk komentar video. Dataset disimpan dalam format CSV dan mencakup kolom teks dan metadata (tanggal, user ID, dsb).

Total data yang dikumpulkan berjumlah Dimana total data yang didapatkan 16134 Komentar, Data disimpan dalam format CSV dan terdiri atas:

1. Teks (komentar atau tweet)
2. Tanggal
3. Username atau user ID (untuk tracking pengguna)
4. Platform (Twitter atau YouTube)

TABEL 1
SAMPLE SUMBER DATASET

Judul Video	Pencarian X yang digunakan
【#holoIDcaf3】 hololive Indonesia 3rd Anniversary Cafe	Holoid dari channel resmi perusahaan hololive indonesia dengan waktu data 1 januari 2023 sampai 01 maret 2025
hololive ID 3rd Anniversary Cafe Teaser #holoIDcaf3	Hololive dari channel resmi perusahaan hololive indonesia dengan waktu data 1 januari 2023 sampai 01 maret 2025
【3D LIVE】 Gebyar hololive Indonesia with all of hololoid! #GebyarholoID	(#KoboKanaeru OR #KaelaKovalskia OR #VestiaZeta OR #MoonaHoshinova OR #Risu OR #Iofifteen OR #AnyaMelfissa OR #Ollie OR #Reine) until:2025-03-01 since:2023-01-01 lang:id)

4.3 Tahapan Preprocessing

Preprocessing dilakukan untuk menghilangkan noise pada data teks sehingga dapat digunakan secara optimal oleh model pembelajaran mesin. Tahapan-tahapan preprocessing meliputi:

4.3.1 Case Folding

Semua huruf dalam data diubah menjadi huruf kecil agar tidak terjadi perbedaan interpretasi antara kata yang sama namun penulisan berbeda.

TABEL 2
HASIL CASE FOLDING

Teks	Case Folding
Manusia manusia lucu banget kelakuannya, ada aja wkwkwwk	manusia manusia lucu banget kelakuannya, ada aja wkwkwwk

4.3.2 Data Cleaning

Menghapus URL, tanda baca, angka, dan karakter khusus.

TABEL 3
HASIL CLEANING

Teks	Cleaning
@@abiekanzy7619 yes. banyak kok yg notes kalo mending kobo sekalian ga ngepost sama sekali, drpd 1 dipost 'normal' dan satunya dipost ngga niat. yg gw dapet dari kasus ini adalah sebisa mungkin perlakuan ke fans	yes banyak kok yg notes kalo mending kobo sekalian ga ngepost sama sekali drpd dipost normal dan satunya dipost ngga niat yg gw dapet dari kasus ini adalah sebisa mungkin perlakuan ke fans

merata, jangan yang satu spesial dan yang satu ngga	merata jangan yang satu spesial dan yang satu ngga
---	--

4.3.3 Data Translate

memastikan konsistensi bahasa dalam dataset dengan menerjemahkan teks ke dalam satu bahasa tertentu.

TABEL 4
TRANSLATE

Teks	Cleaning
wew kobo please comment here u	wew kobo tolong berkomentar di sini u

4.3.4 Tokenizing

Memecah teks menjadi kata-kata.

TABEL 4
HASIL TOKENIZING

Teks	Tokenizing
parasocial relationship itu salah satu senjata content creator terutama streamer lu bisa kayak suisei yang jarang banget service fansnya atau juga kayak nene	['parasocial', 'relationship', 'itu', 'salah', 'satu', 'senjata', 'content', 'creator', 'terutama', 'streamer', 'lu', 'bisa', 'kayak', 'suisei', 'yang', 'jarang', 'banget', 'service', 'fansnya', 'atau', 'juga', 'kayak', 'nene']

4.3.5 Stopword Removal

Menghapus kata-kata umum yang tidak memiliki makna penting.

TABEL 5
HASIL STOPWORD REMOVAL

Teks	Stopword Removal
['parasocial', 'relationship', 'itu', 'salah', 'satu', 'senjata', 'content', 'creator', 'terutama', 'streamer', 'lu', 'bisa', 'kayak', 'suisei', 'yang', 'jarang', 'banget', 'service', 'fansnya', 'atau', 'juga', 'kayak', 'nene']	['parasocial', 'relationship', 'itu', 'salah', 'satu', 'senjata', 'content', 'creator', 'terutama', 'streamer', 'lu', 'bisa', 'kayak', 'suisei', 'yang', 'jarang', 'banget', 'service', 'fansnya', 'atau', 'juga', 'kayak', 'nene']

4.3.6 Stemming

Mengubah kata ke bentuk dasar menggunakan algoritma Sastrawi

TABEL 6
HASIL STEMMING

Teks	Stemming
parasocial relationship	parasocial relationship
salah senjata content	salah senjata content
creator streamer kayak	creator streamer kayak
suisei jarang banget service	suisei jarang banget service
fansnya kayak nen	fansnya kayak nene

4.3.7 Labeling dan Vektorisasi

Labeling dilakukan secara semi-otomatis menggunakan kamus leksikon SenticNet, yang memungkinkan pendekripsi polaritas kata berdasarkan basis pengetahuan afektif. Kata-kata dengan nilai polaritas positif diberi label "positif", sedangkan kata-kata bernilai negatif diberi label "negatif". Proses labeling ini memastikan konsistensi dan memperkuat validitas pengelompokan sentimen. Setelah itu, data dikonversi ke format numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency) agar dapat diproses oleh algoritma klasifikasi.

TABEL 7
SAMPLE LABELING

Positif Label	Negatif Label
I love Oumae Kumiko I	Apakah Amerika (Hololive)
love Oumae Kumiko I	di dunia pervtuberan akan
love Oumae Kumiko I	jatuh seperti Soviet
love Oumae Kumiko I	(NijiSanji)?
love Oumae Kumiko I	
Udah di bahas sama moona, mau jadi idol atau streamer atau dua-duanya terserah bebas, dan ambil projek juga bebas milih	Gen 3 dengan member gimmick nya, Jadi inget rushia dengan segala gimmick GFE nya
masuk akal sih holo serius rambah musik tambah pengaruh virtual idol negara belah tembus panggung gede bentuk virtual self produce lagu koreo nya bikin virtual	aku kecewa ma fansnya yg udah lama malah memperkeruh... pdhal mereka harusnya udah tau grup fb ada yg tiap hari kerjaannya ngegoreng kobo dan memang mereka nunggu

idol beda streamer rubah arah idol kayak berat	santapan ini buat di screenshot
--	---------------------------------

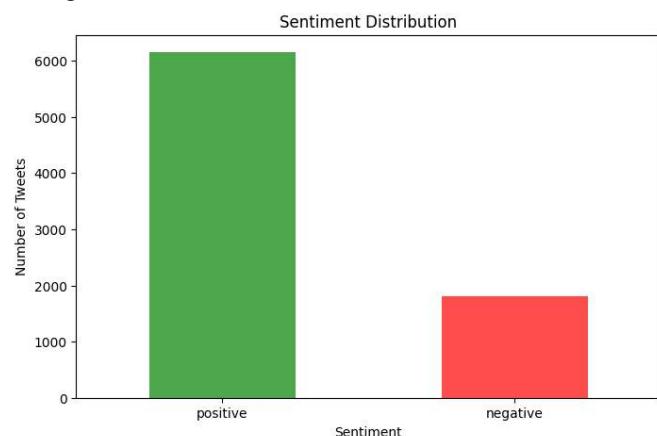
4.4 Distribusi Sentimen

Distribusi sentimen menunjukkan bahwa mayoritas tweet pengguna berada pada kategori positif, diikuti oleh negatif. Kategori netral awalnya terdeteksi, namun telah dihilangkan dari dataset akhir agar klasifikasi lebih fokus ke dua kutub opini utama. Distribusi akhir yang digunakan dalam pelatihan model adalah sebagai berikut:

- Positif: 6152 tweet
- Negatif: 1813 tweet

Distribusi ini menunjukkan bahwa lebih dari 75% opini bersifat positif, yang mengindikasikan penerimaan yang sangat baik terhadap konten dan figur VTuber dari Hololive Indonesia. Namun, ketidakseimbangan jumlah antar kelas juga menjadi perhatian penting yang diselesaikan melalui teknik SMOTE.

Visualisasi distribusi sentimen dapat dilihat dalam grafik batang berikut:



GAMBAR 2
SENTIMENT DISTRIBUTION

4.5 Pembagian Dataset dan Validasi

Dataset dibagi menjadi:

- 80% data latih
- 20% data uji

Juga dilakukan validasi menggunakan 10-fold cross-validation untuk memeriksa konsistensi model.

4.6 Evaluasi Model Tanpa SMOTE

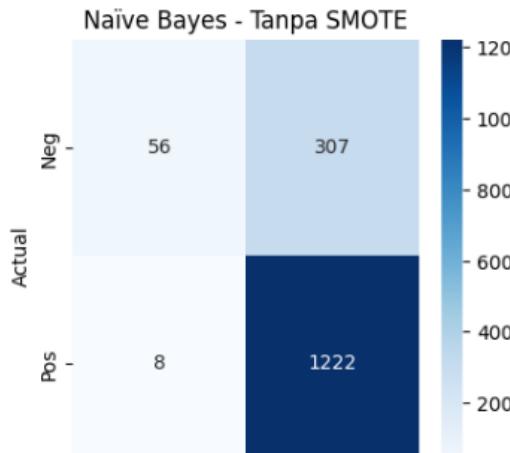
4.6.1 Naïve Bayes

TABEL 8
HASIL NAÏVE BAYES NO SMOTE

Accuracy	0.8022598870056498	
Precision	0.875	0.79921517
Recall	0.15426997	0.99349593
F1 Score	0.26229508	0.8858282

Berdasarkan evaluasi:

- Precision tinggi, menunjukkan model cukup baik dalam mengenali sentimen positif.
- Recall rendah pada kelas negatif menunjukkan kelemahan model dalam menangani kelas minoritas.
- F1-Score menunjukkan ketidakseimbangan performa antar kelas.
- Akurasi cukup baik, namun belum optimal untuk klasifikasi dua kelas.



CONFUSION MATRIX NAÏVE BAYES NO SMOTE

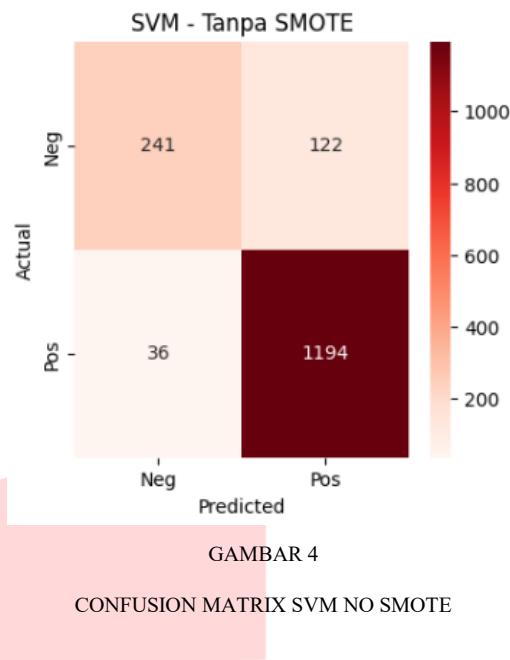
4.6.2 SVM

TABEL 9
HASIL SVM NO SMOTE

Accuracy	0.9008160703075957	
Precision	0.8700361	0.90729483
Recall	0.66391185	0.97073171
F1 Score	0.753125	0.93794187

Kesimpulan dari tabel klasifikasi:

- Precision tinggi untuk kedua kelas mengindikasikan bahwa prediksi positif dan negatif cenderung akurat.
- Recall menunjukkan bahwa model SVM memiliki sensitivitas yang baik, terutama terhadap kelas negatif.
- F1-Score tinggi dan stabil menunjukkan performa yang sangat baik dalam dua kelas.
- Akurasi 80% menjadikan SVM lebih unggul dibandingkan Naïve Bayes dalam klasifikasi sentimen publik.



CONFUSION MATRIX SVM NO SMOTE

4.7 Evaluasi Model dengan SMOTE

4.7.1 Naïve Bayes (dengan SMOTE)

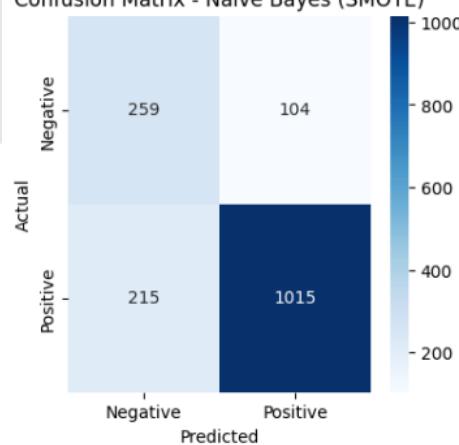
TABEL 10
NAÏVE BAYES SMOTE HASIL

Accuracy	0.7997489014438167	
Precision	0.5464135	0.90705987
Recall	0.71349862	0.82520325
F1 Score	0.61887694	0.86419753

Berdasarkan hasil di atas:

- Penerapan SMOTE memperbaiki keseimbangan recall antar kelas.
- Precision juga menunjukkan hasil yang stabil, menandakan peningkatan pemahaman model terhadap pola minoritas.
- F1-score untuk dua kelas menjadi lebih seimbang.
- Akurasi sedikit meningkat dari sebelumnya.

Confusion Matrix - Naïve Bayes (SMOTE)



GAMBAR 5
CONFUSION MATRIX NAÏVE BAYES SMOTE

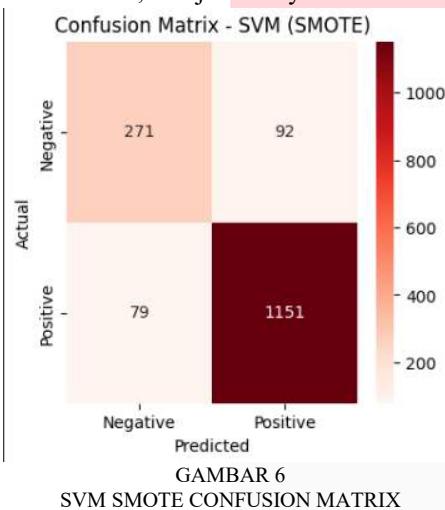
4.7.2 SVM (dengan SMOTE)

TABEL 11
SVM SMOTE HASIL

Accuracy	0.8926553672316384	
Precision	0.77428571	0.92598552
Recall	0.74655647	0.93577236
F1 Score	0.7601683	0.93085321

Berdasarkan evaluasi:

1. Precision tetap tinggi, menunjukkan ketepatan model terhadap dua kategori.
2. Recall meningkat pada kelas positif dibandingkan tanpa SMOTE.
3. F1-Score sangat seimbang, memperkuat efektivitas klasifikasi.
4. Akurasi tertinggi diperoleh oleh kombinasi SVM dan SMOTE, menjadikannya model terbaik.



4.8 Evaluasi K-Fold Cross Validation

Untuk mengevaluasi kestabilan performa model terhadap variasi data latih dan uji, digunakan K-Fold Cross Validation dengan K=10. Menurut Kohavi (1995), 10-fold cross validation memberikan kompromi terbaik antara bias dan variansi.

TABEL 12

K-FOLD HASIL

Model	Rata-Rata Akurasi	Rata-Rata Precision	Rata-rata Recall	Rata-rata F1-Score
Naïve Bayes	0.8025	0.7304	0.7741	0.7453
Smote				
SVM	0.9012	0.8587	0.8619	0.8602
Smote				

Naïve Bayes	0.8053	0.8469	0.5793	0.5821
NoSmote				
SVM	0.9040	0.8864	0.8297	0.8530
NoSmote				

Interpretasi:

1. SVM memiliki performa konsisten lebih tinggi di seluruh metrik rata-rata dibanding Naïve Bayes.
2. Selisih kecil antar fold menunjukkan stabilitas model cukup baik.
3. Evaluasi ini memperkuat kesimpulan bahwa kombinasi SVM + SMOTE merupakan strategi paling optimal dalam penelitian ini.

4.9 Pembahasan Umum

Berdasarkan seluruh hasil eksperimen:

1. SVM dengan SMOTE memberikan hasil terbaik dari seluruh kombinasi model.
2. SMOTE berhasil mengatasi ketidakseimbangan kelas, terutama dalam meningkatkan recall kelas minoritas.
3. Naïve Bayes cocok untuk data sederhana, tetapi performanya tertinggal dibandingkan SVM.
4. Evaluasi metrik seperti precision, recall, dan F1-score memberikan gambaran menyeluruh terhadap kekuatan dan kelemahan masing-masing model.
5. Temuan ini memperkuat landasan teori di Bab II bahwa SVM cocok untuk klasifikasi dengan dua kelas yang jelas, serta bahwa SMOTE dapat membantu memperbaiki ketimpangan kelas.
6. Selain akurasi, penggunaan macro dan weighted average memberikan informasi tambahan mengenai ketimpangan dan performa keseluruhan klasifikasi.

Dengan pendekatan ini, klasifikasi sentimen terhadap VTuber Hololive Indonesia dapat dilakukan secara efektif dan relevan dengan kebutuhan industri hiburan digital yang terus berkembang.

V. KESIMPULAN

Berdasarkan hasil analisis dan pengujian, dapat disimpulkan bahwa akurasi klasifikasi sentimen VTuber Hololive Indonesia menunjukkan variasi performa, di mana model SVM tanpa SMOTE mencapai akurasi tertinggi sebesar 0.8980 pada pembagian data 80:20 dan 0.9040 pada rata-rata 10-Fold Cross Validation. Sebagai perbandingan, akurasi tertinggi yang dicapai model Naïve Bayes adalah 0.8022 (tanpa SMOTE) dan 0.8053 (rata-rata cross-validation tanpa SMOTE). Dengan demikian, SVM, khususnya tanpa penerapan SMOTE, terbukti menjadi model terbaik dalam penelitian ini, menunjukkan keunggulan yang robust dan andal dalam metrik akurasi, presisi,,

REFERENSI

- [1] Z. A. Sadiq, A. Susanti, and Muhammadiyah, "Analisis Media Siber Pada Siaran Langsung Youtuber Ayunda Risu," no. 49, 2022.
- [2] O. N. Tambunan, "Factors in the Growth of Vtuber Livestreaming Entertainment Phenomenon in Japanese Society," vol. 2, no. 1, 2023.
- [3] Akdeas Oktanae Widodo, F. Septiadi, and Nur Aini Rakhmawati, "Analisis Tren Konten Pada Vtuber Indonesia Menggunakan Latent Dirichlet Allocation," *J. Inform. dan Rekayasa Elektron.*, vol. 6, no. 1, pp. 56–63, 2023, doi: 10.36595/jire.v6i1.718.
- [4] D. R. Puspitaningrum and A. Prasetio, "Fenomena 'Virtual Youtuber' Kizuna Ai di Kalangan Penggemar Budaya Populer Jepang di Indonesia," *Mediat. J. Komun.*, vol. 12, no. 2, pp. 128–140, 2019, doi: 10.29313/mediator.v12i2.4758.
- [5] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [6] D. Purnamasari *et al.*, *Pengantar Metode Analisis Sentimen*. 2023.
- [7] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [8] W. Widayat, "Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1018, 2021, doi: 10.30865/mib.v5i3.3111.
- [9] M. Reyus Liman Prodi Perhotelan and S. Yapari Bandung, "Pengaruh Hololive Indonesia dalam Memperkenalkan Bahasa Indonesia Kepada Penonton Internasional," vol. 1, no. 2, 2023.
- [10] J. I. Adzania and A. Arianingsih, "Dampak Menonton Vtuber Terhadap Pengemar Budaya Populer Jepang Di Bandung," *MAHADAYA J. Bahasa, Sastra, dan Budaya*, vol. 2, no. 2, pp. 267–278, 2022, doi: 10.34010/mhd.v2i2.8265.
- [11] R. Nanda, E. Haerani, S. K. Gusti, and S. Ramadhani, "Klasifikasi Berita Menggunakan Metode Support Vector Machine," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 2, pp. 269–278, 2022, doi: 10.32672/jnkti.v5i2.4193.
- [12] Y. A. Singgalen, "Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Sistematic Literature Review," *J. Inf. Syst. Informatics*, vol. 3, no. 2, pp. 278–302, 2021, doi: 10.33557/journalisi.v3i2.125.
- [13] I. Daqiqil, H. Saputra, Syamsudhuha, R. Kurniawan, and Y. Andriyani, "Sentiment analysis of student evaluation feedback using transformer-based language models," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 36, no. 2, pp. 1127–1139, 2024, doi: 10.11591/ijeecs.v36i2.pp1127-1139.
- [14] A. Aulia, M. R. Tanjung, M. Iqbal, and R. F. Wijaya, "Implementasi Algoritma Naïve Bayes Dalam Menganalisis Jumlah Live Stream VTuber Skem," *Bull. Inf. Technol.*, vol. 4, no. 4, pp. 554–559, 2023, doi: 10.47065/bit.v4i4.899.
- [15] S. Informasi, U. Darwan Ali, J. Batu Berlian, and K. Tengah, "Penerapan Metode Naive Bayes Dalam Mengklasifikasi Penerima BLT Pada Desa Pelangsian," *J. JUPITER*, vol. 14, no. 2, pp. 64–70, 2022.
- [16] K. V. S. Toy, Y. A. Sari, and I. Cholissodin, "Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5068–5074, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>