

Klasifikasi Komentar *YouTube* pada *Video* Pembelajaran Bahasa Inggris Menggunakan *Naïve Bayes Classifier* (NBC)

Zaehol Fatah¹, Eny Itsnainy Agustin²

^{1,2} Universitas Ibrahimy, Situbondo

Email: itsnainyeny333@gmail.com

Abstrak

Perkembangan teknologi informasi yang pesat telah menempatkan platform *YouTube* sebagai sumber utama untuk memperoleh informasi dan pendidikan, khususnya dalam konteks pembelajaran bahasa Inggris, yang menghasilkan volume umpan balik pengguna yang besar dalam bentuk komentar. Tujuan penelitian ini adalah menganalisis dan mengklasifikasikan sentimen (Positif, Negatif, dan Netral) dari 50.954 komentar pada *video* pembelajaran bahasa Inggris di *YouTube* dengan mengaplikasikan algoritma *Naïve Bayes Classifier* (NBC). Metode penelitian kuantitatif ini melibatkan tahapan *text mining* yang komprehensif, mulai dari preprocessing data, pelabelan otomatis menggunakan *VADER Sentiment Analyzer*, hingga pembentukan matriks fitur (*confusion matrix*) dengan TF-IDF. Hasil evaluasi kinerja model NBC menunjukkan Akurasi Keseluruhan sebesar 84%. Walaupun akurasi tinggi, ditemukan adanya ketidakseimbangan kelas (*class imbalance*) yang signifikan, dengan sentimen Netral mendominasi (56.34%). Analisis metrik per kelas mengungkapkan bahwa model lemah dalam mengklasifikasikan sentimen Negatif (memiliki Recall 0.41), dengan 313 komentar Negatif salah diprediksi sebagai Positif. Penelitian menyimpulkan bahwa NBC merupakan solusi valid dan efektif, namun disarankan untuk mengaplikasikan teknik penanganan *class imbalance* pada studi mendatang guna meningkatkan kemampuan deteksi sentimen Negatif, yang merupakan umpan balik penting untuk peningkatan kualitas konten.

Kata Kunci: Analisis Sentimen, *Naïve Bayes Classifier*, *YouTube*, Pembelajaran Bahasa Inggris, *Class Imbalance*

Abstract

The rapid development of information technology has positioned YouTube as a primary platform for accessing information and education, particularly within the context of English language learning, which generates a substantial volume of user feedback in the form of comments. This research aims to analyze and classify the sentiments (Positive, Negative, and Neutral) of 50,954 comments on English learning videos on YouTube using the Naïve Bayes Classifier (NBC) algorithm. This quantitative research employs a comprehensive text mining workflow, including data preprocessing, automated labeling using the VADER Sentiment Analyzer, and feature matrix construction with TF-IDF. The performance evaluation of the NBC model shows an overall accuracy of 84%. Despite this high accuracy, a significant class imbalance was identified, with Neutral sentiment dominating (56.34%). Class-wise metric analysis reveals that the model underperforms in classifying Negative sentiment (with a Recall of 0.41), where 313 Negative comments were misclassified as Positive. This study concludes that NBC serves as a valid and effective approach; however, it is recommended that future research incorporate class imbalance handling techniques to enhance the detection of Negative sentiment, which provides essential feedback for improving content quality.

Keywords: *Class Imbalance, English Language Learning, Naïve Bayes Classifier, Sentiment Analysis, YouTube*

PENDAHULUAN

Perkembangan dalam bidang teknologi informasi telah menyebabkan transformasi signifikan pada cara individu mendapatkan dan memanfaatkan data. Salah satu *platform online* yang berkembang dengan sangat cepat adalah *YouTube*. *YouTube* berfungsi sebagai mesin pencari terbesar kedua setelah *Google* dan merupakan salah satu *platform* berbagi *video* terpenting di seluruh dunia. *Platform* ini memberikan kesempatan bagi penggunaannya untuk mengunggah, menonton, berinteraksi, serta memberikan umpan balik pada konten *video*. Satu miliar jam *video* lebih ditonton setiap harinya oleh dua miliar pengguna aktif *YouTube* (Jamil et al., 2024). Pada zaman ini, *YouTube* telah berkembang dan berperan sebagai salah satu *platform* utama untuk memperoleh informasi, hiburan, dan pendidikan di tingkat global. Dalam dunia pendidikan, *YouTube* banyak digunakan sebagai alat bantu pembelajaran karena mampu menyajikan konten secara visual, inovatif, interaktif, dan efisien. Selain itu, *platform* ini mudah diakses serta dianggap menarik oleh siswa atau peserta didik dari berbagai latar belakang (Fatah, 2025; Mareta et al., 2025; Nur Hidayah, 2021).

Salah satu sektor pendidikan yang sangat memanfaatkan *platform YouTube* adalah pembelajaran bahasa Inggris. Dengan semakin kuatnya pengaruh globalisasi, masyarakat dituntut untuk menyesuaikan diri. Kondisi ini mendorong banyak orang untuk mengembangkan kemampuan bahasa, khususnya Bahasa Inggris, dikarenakan bahasa Inggris telah menjadi bahasa dunia yang memiliki peran signifikan di berbagai disiplin ilmu sejak ditetapkannya sebagai bahasa internasional (Nur Hidayah, 2021). Seiring meningkatnya jumlah *video* pembelajaran bahasa Inggris yang diunggah, semakin banyak pula pengguna *internet* yang memberikan tanggapan terhadap *video* yang dipublikasikan. Hal ini terlihat dari beragam komentar pada *video* pembelajaran bahasa Inggris. Dari berbagai

pandangan tersebut, dilakukan pengelompokan menjadi pendapat atau kecenderungan yang bersifat positif maupun negatif. Proses pengelompokan ini dikenal sebagai analisis sentimen (Febrialdi Ansyah et al., 2024).

Pengelompokan komentar atau sentimen dalam studi ini dilakukan dengan memanfaatkan algoritma *Naïve Bayes Classifier* (NBC). NBC adalah algoritma yang mudah dipahami, memiliki kecepatan perhitungan yang tinggi, memerlukan jumlah data latih yang relatif sedikit, serta memiliki tingkat efektivitas yang baik (Morgan et al., 2022; Rahman et al., 2020). Penelitian terkait analisis sentimen telah dilakukan oleh (Yasir & Suraji, 2023). Temuan dari penelitian tersebut menunjukkan tingkat ketepatan (*accuracy*) sebesar 90% menggunakan algoritma *Naïve Bayes*, 83% menggunakan algoritma *Decision Tree*, dan 87% menggunakan algoritma *Random Forest*. Hasil tersebut memberi tahu bahwa algoritma NBC memiliki kinerja yang lebih bagus dalam pengklasifikasian sentimen dibandingkan dengan algoritma lainnya.

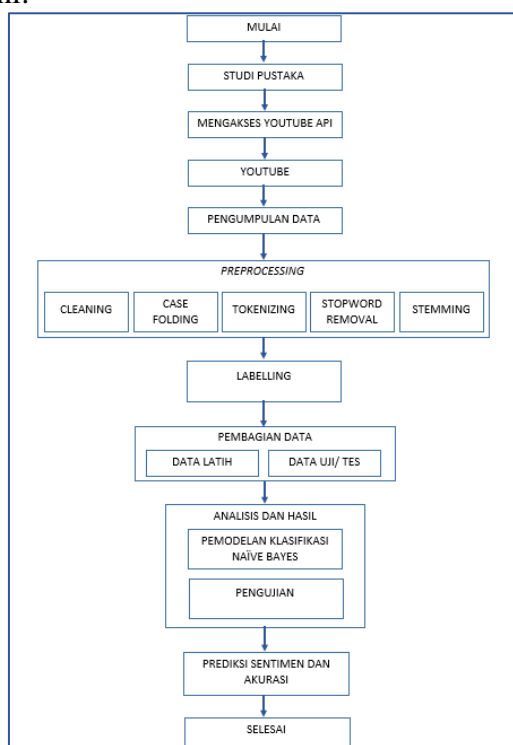
Dengan mempertimbangkan latar belakang tersebut, penelitian memiliki tujuan yaitu mengevaluasi sentimen melalui penerapan algoritma *Naïve Bayes Classifier* dalam mengklasifikasikan komentar pada *video* pembelajaran bahasa Inggris di *platform YouTube*, sehingga para pembuat konten dapat memaksimalkan kualitas *video* pembelajaran bahasa Inggris pada saluran *YouTube* mereka.

METODE

Penelitian yang mengimplementasikan metode analisis sentimen berbasis *text mining* ini mengaplikasikan algoritma *Naïve Bayes Classifier* (NBC) untuk mengetahui lebih banyak komentar positif, negatif atau netral yang terlontar di kolom komentar *video* pembelajaran Bahasa Inggris. Proses penelitian dilakukan melalui beberapa tahapan utama, yaitu studi pustaka, pengumpulan data komentar, pra-

pemrosesan data (*preprocessing*), pelabelan data serta pelatihan dan pemodelan klasifikasi menggunakan *Naïve Bayes Classifier* (NBC) (Al et al., 2022; Aqsalia et al., 2024; Febrialdi Ansyah et al., 2024; Huwaida et al., 2024).

Gambar 1 merupakan gambaran alur proses yang dilalui dalam penelitian ini:



Gambar 1 Tahapan Klasifikasi Data Menggunakan NBC

Berdasarkan Gambar (1), berikut adalah penjelasan dari alur proses penelitian:

1. Studi Pustaka

Studi pustaka merupakan tahap mencari dan menganalisis karya ilmiah yang terkait dengan masalah yang akan diteliti. Sasaran dari kegiatan ini adalah mengidentifikasi teori-teori yang berguna dalam menjawab masalah penelitian dan sebagai landasan untuk melakukan penelitian (Aqsalia et al., 2024).

2. Penhimpunan Data (Mengakses YouTube dan YouTube API)

Data komentar diperoleh dari *video* pembelajaran bahasa Inggris pada *platform YouTube* yang memiliki tingkat interaksi tinggi yaitu *video* dengan judul “ONE language, THREE accents - UK vs. USA vs. AUS English! (+ Free PDF)” di *channel YouTube* English with Lucy. Komentar dikumpulkan menggunakan layanan/fasilitas yang telah disediakan oleh *Platform YouTube* sendiri yakni *YouTube Data API v3* (Rahman et al., 2020; Taufiqurahman et al., 2024). Data yang didapat adalah sebanyak 51.839 komentar dan dibersihkan melalui proses *cleaning data* sehingga diperoleh 50.954 data komentar yang layak diproses ke tahap selanjutnya. Seluruh data yang telah dibersihkan kemudian melalui proses *praproses (preprocessing)*, meliputi pembersihan tanda baca, simbol, angka, *case folding*, *tokenization*, *stopword removal*, dan *stemming*.

3. Preprocessing

Bagian ini merupakan proses pengubahan data komentar yang masih tidak terorganisasi dengan baik agar berubah menjadi data yang siap dan layak untuk dianalisis (Huwaida et al., 2024). Seluruh rangkaian proses *preprocessing* pada penelitian ini dilakukan menggunakan Python versi 3.10.6. Adapun tahapan *preprocessing* yang digunakan meliputi:

Cleaning: Simbol, emoji, tanda baca, tautan URL, angka, dan karakter lainnya dibersihkan dari data set ini. Pada tahap ini pula baris komentar yang tidak jelas (seperti: hanya berisi simbol) dihapus. Tahap *cleaning* seperti pada Tabel (1).

Tabel 1 Proses Cleaning

Input	Output
"14:16 > my mom says sandals"	my mom says sandals
"How do you say to highway/freeway/interstate in British English ?"	How do you say to highway freewayinterstate in British English
"The way she says audible sounds sweet & amp"	The way she says audible sounds sweet amp
"Thank you Lucy for sharing these differences in the video."	Thank you Lucy for sharing these differences in the video

Case Folding: Dataset yang berisi huruf kapital diganti dengan huruf non-kapital. Tahap case folding seperti pada Tabel (2).

Tabel 2 Proses Case Folding

Input	Output
my mom says sandals	my mom says sandals
How do you say to highway freewayinterstate in British English	how do you say to highway freewayinterstate in british english
The way she says audible sounds sweet amp	the way she says audible sounds sweet amp
Thank you Lucy for sharing these differences in the video	thank you lucy for sharing these differences in the video

Tokenizing: Kalimat komentar dipisahkan menjadi kata-kata individual. Hasil dari tahap *tokenizing* sendiri tidak dapat terlihat dengan mata telanjang karena komentar sudah bersih dari semua jenis tanda baca ketika proses *cleaning* sehingga *outputnya* hampir mirip dengan hasil proses *lowercase*. Tahap *tokenizing* seperti pada Tabel (3).

Tabel 3 Proses Tokenizing

Input	Output
my mom says sandals	my mom says sandals
how do you say to highway freewayinterstate in british english	how do you say to highway freewayinterstate in british english
the way she says audible sounds sweet amp	the way she says audible sounds sweet amp
thank you lucy for sharing these differences in the video	thank you lucy for sharing these differences in the video

Stopword Removal: Menghilangkan kata sambung atau unsur-unsur bahasa lain yang tidak memiliki pengaruh signifikan terhadap makna kalimat. Tahap *stopword removal* seperti pada Tabel (4).

Tabel 4 Langkah Penghilangan Stopword

Input	Output
my mom says sandals	mom says sandals
how do you say to highway freewayinterstate in british english	say highway freewayinterstate british english

the way she says audible sounds sweet amp	way says audible sounds sweet amp
thank you lucy for sharing these differences in the video	thank lucy sharing differences video

Stemming: Dalam *step* ini, imbuhan yang ada pada kata-kata dibuang sehingga menjadi kata dasar (Al et al., 2022; Huwaida et al., 2024). Tahap *stemming* seperti pada Tabel (5).

Tabel 5 Proses Stemming

Input	Output
mom says sandals	mom says sandals
say highway freewayinterstate british english	say highway freewayinterstate british english
way says audible sounds sweet amp	way says audible sounds sweet amp
thank lucy sharing differences video	thank lucy sharing differences video

4. Labelling

Labelling merupakan langkah untuk memberikan tanda pada setiap komentar. Tahap *labelling* dilaksanakan dengan menghitung keseringan (*frequency*) kata yang muncul untuk diberikan penanda positif, negatif atau netral pada data komentar yang sudah melewati tahap *preprocessing* (Morgan et al., 2022; Rahman et al., 2020). Karena volume data yang sangat besar (lebih dari 50.000 baris), pelabelan manual tidak memungkinkan dilakukan. Oleh karena itu, penelitian ini menggunakan pendekatan *labelling* otomatis (*Preprocess-and-Label*). Proses ini dilakukan pada data teks yang telah dibersihkan (setelah *stemming*) agar hasil pelabelan menjadi lebih akurat. Untuk pelabelan, penelitian ini menerapkan VADER Sentiment Analyzer karena memiliki tingkat ketepatan (*accuracy*) yang lebih baik dibandingkan dengan alat lain seperti TextBlob. VADER merupakan metode berbasis leksikon yang sesuai digunakan untuk menganalisis data dari media sosial karena sistem ini memberikan skor sentimen pada setiap kata maupun frasa. Dengan begitu, VADER dapat mengidentifikasi suatu teks yang bersifat positif (1), negatif (0), atau netral (2). Dalam prosesnya, VADER melihat keseluruhan isi teks

dengan memperhitungkan berbagai elemen, seperti penekanan, penggunaan emotikon, kata negasi, serta kata-kata yang berfungsi menguatkan makna. Kategori sentimen negatif, positif, maupun netral ditentukan melalui nilai *compound* yang berada pada rentang -1 (sangat negatif) hingga $+1$ (sangat positif) (Anam, 2025; Arrahman Firsttama et al., 2024). Tahap *labelling* seperti pada Tabel (6).

Tabel 6 Proses Labelling

Input	Output
mom says sandals	2
say highway freewayinterstate british english	2
way says audible sounds sweet amp	1
thank lucy sharing differences video	1

5. Penerapan Feature Extraction (TF-IDF) dan Pembagian Data

TF-IDF (*Term Frequency–Inverse Document Frequency*) termasuk metode statistik yang berfungsi untuk menilai tingkat kepentingan unit kata dalam sebuah dokumen yang diukur melalui jumlah kemunculannya di dokumen tersebut serta tingkat kelangkaannya di keseluruhan kumpulan dokumen (korpus). Tahap ini berfungsi sebagai jembatan antara data teks yang sudah bersih (hasil

preprocessing) dengan model klasifikasi matematis (Naïve Bayes) karena data harus diubah menjadi matriks angka. Tujuan Utama dari TF-IDF yaitu menghasilkan Matriks Fitur (X_{matrix}), yang mengatur setiap baris sebagai satu komentar dan setiap kolom sebagai token atau kata unik. Matriks ini, bersama dengan Vektor Label Sentimen (Y) yang berisi kategori 0, 1, dan 2 (Negatif, Positif, Netral), menjadi input utama bagi model. Bobot di dalam matriks X ditentukan berdasarkan keseringan (*frequency*) kemunculan kata dalam satu dokumen (TF) dikalikan dengan kebalikan dari frekuensi kemunculan kata di seluruh dokumen (IDF). Rumus ini memberikan bobot yang tinggi untuk kata-kata yang sering muncul dalam sebuah dokumen namun jarang muncul di dokumen lain, sehingga efektif dalam mengekstrak fitur penting untuk teks (Misrun et al., 2024). Berikut adalah rumus menghitung *TF-IDF*:

$$TF - IDF(t, d) = \frac{TF(t, d)}{\times IDF(t)}$$

dengan:

- $TF(t, d)$ adalah Term Frequency, yaitu frekuensi kemunculan term t dalam dokumen d , biasanya dihitung sebagai:

$$TF(t, d) = \frac{\text{jumlah kemunculan term } t \text{ dalam dokumen}}{\text{total kata dalam dokumen } d}$$

- $IDF(t)$ (*Inverse Document Frequency*) adalah nilai yang dipakai untuk menentukan tingkat signifikansi suatu term t dalam seluruh dokumen, dan besarnya ditentukan melalui perhitungan sebagai berikut:

$$IDF(t) = \log\left(\frac{N}{DF(t)}\right)$$

di mana N menyatakan jumlah keseluruhan dokumen pada korpus,

sedangkan $DF(t)$ menggambarkan berapa banyak dokumen yang mengandung *term* tersebut. Setelah **Matriks Fitur (X)** dan **Vektor Label Sentimen (Y)** berhasil dikonversi ke format yang diperlukan, langkah berikutnya adalah melakukan proses *Train/Test Split*. Pada tahap ini, kedua komponen tersebut dipisahkan menjadi dua subset, yaitu **set pelatihan (training set)** dan **set pengujian (testing set)**. Subset untuk pelatihan berfungsi sebagai dasar dalam membangun dan menyesuaikan **model klasifikasi**, sementara subset untuk pengujian digunakan guna menilai kualitas dan kemampuan model dalam memprediksi data baru. Pembagian antara *training set* dan *testing set* biasanya mengikuti dua skenario umum, yakni porsi *training* lebih kecil dibandingkan *testing*, atau sebaliknya, bagian *training* disusun dengan jumlah yang lebih besar dari *testing* (Aqsalia et al., 2024). Pembagian dilakukan menggunakan rasio standar (umumnya 8 : 2, 8 untuk Data *train* dan 2 Data *test*). Dalam klasifikasi multi-kelas, metode *Stratified Sampling* digunakan saat pembagian untuk memastikan proporsi ketiga kelas sentimen (Positif, Negatif, Netral) tetap terjaga di kedua set, sehingga hasil evaluasi model menjadi lebih representatif.

6. Naïve Bayes Classifier

Algoritma NBC adalah teknik untuk menghitung peluang suatu data yang diuji untuk masuk ke dalam kategori spesifik didasarkan pada kemungkinan dari setiap atribut data yang diuji tersebut (Aqsalia et al., 2024). Teorema Bayes adalah prinsip yang digunakan untuk meramalkan kejadian di masa depan dengan mengacu pada pengalaman yang telah diperoleh sebelumnya. Salah satu keuntungan dari algoritma naïve bayes classifier adalah kemudahan dalam penerapan dan tingkat akurasi yang tinggi (Misrun et al., 2024). Dalam penelitian

ini, sentimen akan diklasifikasikan menggunakan *Naïve Bayes Classifier* Multinomial ke dalam 3 kategori yakni positif, negatif dan netral. *Multinomial Naïve Bayes* dikembangkan dari algoritma *naïve bayes* yang dirancang khusus untuk menangani tugas klasifikasi berbasis teks maupun dokumen. Model ini menjadi salah satu pendekatan yang paling umum digunakan dalam analisis teks karena mampu memproses data yang direpresentasikan dalam bentuk vektor frekuensi kata. *Multinomial Naïve Bayes* ideal diterapkan ketika setiap fitur menggambarkan jumlah kemunculan suatu elemen, seperti pada pemodelan teks di mana setiap fitur merepresentasikan frekuensi kata dalam dokumen.

$$P(C_i|X) \propto P(C_i) \times \prod_{k=1}^{\{n\}} P(x_k|C_i)$$

Rumus (1) merupakan rumus *Naïve Bayes Classifier* Multinomial dengan asumsi independensi di mana:

- $P(C_i|X)$ merupakan probabilitas terbaru dari kelas C_i terhadap data X , yakni nilai yang menunjukkan seberapa besar kemungkinan data X termasuk dalam kelas C_i ,
- $P(C_i)$ adalah probabilitas awal (prior) dari kelas C_i ,
- $P(x_k|C_i)$ adalah probabilitas kemunculan fitur ke- k (misalnya kata dalam data) dalam kelas C_i ,
- produk dari semua fitur dihitung berdasarkan frekuensi kata-kata atau bobot TF-IDF dalam data tersebut (Huwaida et al., 2024).

Secara lebih detail, rumus ini mengasumsikan distribusi multinomial dan sering digunakan dalam klasifikasi teks dan dokumen, khususnya saat fitur diwakili sebagai frekuensi kata atau bobot TF-IDF. Model ini menghitung peluang setiap kelas berdasarkan frekuensi fitur dan kemudian memilih kelas dengan peluang tertinggi sebagai

hasil prediksi.

7. Prediksi Sentimen dan Akurasi

Prediksi sentimen dilakukan dengan mengklasifikasikan komentar *YouTube* menjadi kategori positif, negatif dan netral menggunakan algoritma NBC berdasarkan fitur yang telah diekstraksi dari data komentar. Akurasi pengujian dihitung untuk mengetahui seberapa tepat model dapat memetakan kategori sentimen komentar sesuai label data sebenarnya. Tingkat akurasi menjadi indikator utama efektivitas metode *Naïve Bayes* dalam menganalisis sentimen pada komentar *video YouTube* (Khanafiyah et al., 2025).

HASIL DAN PEMBAHASAN

Hasil

Hasil penelitian ini disajikan secara bertahap, dimulai dari statistik data pra-pemrosesan (*preprocessing*), dimensi *Feature Extraction*, hingga hasil utama evaluasi kinerja model *Naïve Bayes Classifier*.

Hasil Preprocessing Data

Data yang digunakan merupakan kumpulan komentar *video YouTube* yang telah melalui serangkaian tahapan preprocessing seperti *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* menggunakan *Porter Stemmer*. Jumlah akhir data yang lolos *cleaning* dan siap dianalisis adalah 50.954 komentar.

Tahap pelabelan (*Labelling*) dilakukan secara otomatis menggunakan *VADER Sentiment Analyzer* pada teks yang telah melalui proses *stemming*. Tabel 1 menunjukkan distribusi data akhir berdasarkan label sentimen (0: Negatif, 1: Positif, 2: Netral).

Tabel 7 Distribusi Label Sentimen Data Komentar YouTube

Kategori Sentimen	Label	Jumlah Data/ Dokumen	Persentase
Negatif	0	5.549	10.89%
Positif	1	16.696	32.77%
Netral	2	28.709	56.34%
Total	-	50.954	100.00%

Dari Tabel (7), terlihat adanya ketidakseimbangan kelas (*class imbalance*) yang signifikan, di mana komentar Netral mendominasi (56.34%) dibandingkan Positif (32.77%) dan Negatif (10.89%).

Tabel (8) menunjukkan contoh transformasi teks dari awal hingga akhir tahap *preprocessing* data, memastikan teks sudah terstandarisasi sebelum *Feature Extraction*.

Tabel 8 Tabel Transformasi Teks Hasil Preprocessing

Komentar Asli	Cleaning & Case Folding	Stopword Removed	Stemmed (Final)
"14:16 my mom says sandals"	my mom says sandals	mom says sandals	mom say sandal
"How do you say to highway/freeway/interstate in British English ?"	how do you say to highway freewayinterstate in british english	say highway freewayinterstate british english	say highway freewayinterst british english
"The way she says audible sounds sweet &"	the way she says audible sounds sweet amp	way says audible sounds sweet amp	way say audibl sound sweet amp
"Thank you Lucy for sharing these differences in the video."	thank you lucy for sharing these differences in the video	thank lucy sharing differences video	thank luci share differ video

Hasil Feature Extraction (TF-IDF) dan Pembagian Data

Tabel 9 Hasil Ekstraksi Fitur (TF-IDF) dan Pembagian Data

Deskripsi	Jumlah Baris (Data/Dokumen)	Persentase
Total Data	50.954	100%
Matriks Fitur (X) Dimensi	(50954, 5000)	-
Data Latih (X_{train}, Y_{train})	40.763	80%
Data Uji (X_{test}, Y_{test})	10.191	20%

Proses *Feature Extraction* menggunakan *TF-IDF* diterapkan pada teks hasil *stemming*. Dengan batasan $max_features=5000$ dan $ngram_range=(1, 2)$, dihasilkan Matriks Fitur (X) dengan dimensi 50.954×5.000 , menunjukkan bahwa 5.000 fitur (kata/bigram) paling informatif telah diekstrak. Hasilnya ditunjukkan pada Tabel (9).

Tahap *Train/Test Split* membagi Matriks Fitur menjadi 8 : 2, (80% Data Latih = 40.763 komentar) dan (20% Data Uji = 10.191 komentar). Pembagian ini menggunakan metode *Stratified Sampling* untuk memastikan proporsi sentimen (Positif, Negatif, Netral) yang tidak

seimbang tetap terjaga di kedua set, sehingga model dievaluasi secara adil.

Hasil Evaluasi Model Naïve Bayes Classifier

Model *Naïve Bayes Classifier* (*NBC*) dilatih menggunakan Matriks Fitur dari Data Latih dan dievaluasi kinerjanya menggunakan Data Uji. Kinerja model diukur melalui empat metrik utama: *Accuration*, *Precision*, *Recall*, dan *F1-Score*.

Tabel 10 Hasil Metrik Evaluasi Kinerja Model NBC

Metrik	Negatif (0)	Positif (1)	Netral (2)	Rata-rata (Weighted Avg)
Precision	0.86	0.81	0.86	0.84
Recall	0.41	0.82	0.93	0.84
F1-Score	0.56	0.81	0.89	0.83
Akurasi Keseluruhan	-	-	-	0.84

Secara keseluruhan, model NBC berhasil mencapai Akurasi Keseluruhan sebesar 84% dalam memprediksi sentimen komentar. Angka ini menunjukkan bahwa model dapat mengklasifikasikan label sentimen dengan benar pada 84 dari setiap 100 komentar uji.

Analisis per kelas (Tabel 10) menunjukkan bahwa kelas Netral memiliki performa terbaik, terutama pada metrik *Recall* (0.93) dan *F1-Score* (0.89), menunjukkan model sangat baik dalam mengidentifikasi sebagian besar komentar yang benar-benar Netral. Sebaliknya, kelas Negatif memiliki kinerja terendah (*Recall* 0.41 dan *F1-Score* 0.56), menunjukkan bahwa komentar yang sebenarnya Negatif cenderung sulit diklasifikasikan dengan tepat, sering dikelirukan dengan kelas lain.

Untuk memahami sumber kesalahan prediksi, Matriks Kebingungan (*Confusion Matrix*) disajikan pada Tabel (11).

Tabel 11 Confusion Matrix

	Prediksi Negatif (0)	Prediksi Positif (1)	Prediksi Netral (2)
Aktual Negatif	455	313	34
Aktual Positif	39	2738	56
Aktual Netral	33	348	536

Pembahasan

Bagian ini membahas hasil-hasil penelitian yang telah diperoleh serta menginterpretasikan makna dan implikasinya dalam konteks teori yang relevan dan studi sebelumnya.

Interpretasi Kinerja Model NBC

Akurasi sebesar 84% menunjukkan bahwa algoritma NBC sangat efektif dalam

mengklasifikasikan sentimen komentar video pembelajaran Bahasa Inggris. Kinerja ini didukung oleh dua faktor utama:

1. Ekstraksi Fitur TF-IDF yang Optimal: Penggunaan ngram_range (1, 2) dan pembatasan fitur hingga 5.000 berhasil menangkap konteks sentimen melalui pasangan kata (*bigram*) yang lebih akurat (Jelita & Saad, 2025).
2. Efektivitas *Preprocessing*: Kombinasi *stopword removal* dan *Porter Stemmer* memastikan bahwa Matriks Fitur (X) hanya berisi kata-kata dasar yang relevan dengan sentimen. Proses *stemming* dengan menghilangkan kata melalui beberapa iterasi sebelum semua aturan atau kondisi dipertimbangkan (Polus, 2021) efektif meminimalkan *noise* dan meningkatkan probabilitas dalam perhitungan *Naïve Bayes*.

Kinerja yang baik ini juga terkonfirmasi melalui nilai *Weighted Average* (Rata-rata Tertimbang) untuk *Precision* dan *Recall* yang mencapai 0.84, serta *F1-Score* 0.83.

Analisis Berdasarkan Metrik Per Kelas dan *Confusion Matrix*

Evaluasi kinerja model klasifikasi sentimen pada komentar *YouTube* dilakukan dengan memanfaatkan sejumlah metrik, yaitu akurasi, presisi, *recall*, *F1-score*, serta *confusion matrix* (Putri et al., 2025).

Meskipun akurasi keseluruhan tinggi, analisis per kelas (Tabel 10) mengungkap adanya variasi performa, terutama pada kelas Negatif.

Kelas Netral (2) mencapai performa paling optimal dengan metrik *Recall* paling tinggi (0.93) dan *F1-Score* (0.89). Hal ini membuktikan bahwa model sangat andal dalam mengenali komentar yang sebenarnya Netral. Sebaliknya, kelas Negatif (0) memiliki *Recall* terendah, hanya 0.41, meskipun memiliki *Precision* yang tinggi (0.86).

Rendahnya *Recall* pada kelas Negatif ini disebabkan oleh dua faktor utama yang terlihat dari *Confusion Matrix* (Tabel 11) dan distribusi data:

1. Isu *Class Imbalance*: Kelas Negatif adalah kelas minoritas ekstrem (hanya 10.89% dari total data, atau sekitar 1.110 sampel dalam data uji). Karena *Naïve Bayes* menggunakan probabilitas prior kelas, model cenderung bias untuk memprediksi ke kelas mayoritas (Netral dan Positif) untuk meminimalkan kesalahan secara keseluruhan.
2. Kesalahan Klasifikasi: Berdasarkan *Confusion Matrix* (Tabel 11), dari total 802 komentar Aktual Negatif:
 - Hanya 455 yang diklasifikasikan dengan benar (True Positive).
 - Sebanyak 313 komentar Negatif salah diklasifikasikan sebagai Positif.
 - Sebanyak 34 komentar Negatif salah diklasifikasikan sebagai Netral.

Jumlah *False Negative* yang besar (347 kasus, terutama dikelirukan sebagai Positif) inilah yang secara langsung menekan nilai *Recall* kelas Negatif. Hal ini mungkin terjadi karena komentar Negatif seringkali menggunakan bahasa yang ambigu, sarkasme, atau frasa yang tidak terdeteksi secara eksplisit sebagai negatif oleh *VADER* atau bobot *TF-IDF*, sehingga model menganggapnya sebagai Positif atau Netral.

Meskipun demikian, Precision yang tinggi (0.86) pada kelas Negatif mengindikasikan bahwa ketika model berani memprediksi suatu komentar sebagai Negatif, prediksi tersebut kemungkinan besar benar. Fenomena ini sejalan dengan sifat *Naïve Bayes* yang bekerja sangat baik pada klasifikasi teks dengan asumsi independensi fitur (Rahman et al., 2020).

Perbandingan dengan Penelitian Sebelumnya

Hasil Akurasi 84% dari studi ini membuktikan bahwa NBC adalah algoritma yang efisien dan efektif untuk klasifikasi sentimen pada komentar video pembelajaran Bahasa Inggris.

Kinerja ini sebanding dan berada dalam rentang hasil penelitian sentimen berbasis *Naïve Bayes* sebelumnya. Misalnya, penelitian oleh (Yasir & Suraji, 2023) yang menggunakan *Naïve Bayes* untuk analisis sentimen di *YouTube* juga mencapai tingkat akurasi yang tinggi, yaitu 90%. Perbedaan Akurasi antara 84% dalam penelitian ini dan 90% dalam studi tersebut dapat dijelaskan oleh beberapa faktor metodologis:

- Topik Data dan Bahasa: Perbedaan topik (*Video Pembelajaran Bahasa Inggris* vs. Topik lainnya) dan audiens dapat memengaruhi kompleksitas bahasa, penggunaan kosakata, dan tingkat *noise* dalam komentar.
- Perbedaan *Dataset* dan *Labelling*: Ukuran dan kualitas dataset serta metode pelabelan (meskipun kedua studi menggunakan *Naïve Bayes*) akan memengaruhi hasil akhir.
- Parameter *TF-IDF*: Parameter *max_features* dan *ngram_range* yang digunakan dalam penelitian ini mungkin berbeda, yang secara langsung memengaruhi kualitas Matriks Fitur yang diumpangkan ke NBC.

Secara keseluruhan, *Naïve Bayes Classifier* teridentifikasi sebagai solusi yang valid untuk analisis sentimen, namun untuk meningkatkan *Recall* pada kelas minoritas (Negatif), perlu dipertimbangkan teknik penanganan *class imbalance* (*oversampling* atau *undersampling*) pada penelitian selanjutnya.

SIMPULAN (PENUTUP)

Penelitian ini ditujukan untuk mengidentifikasi dan mengklasifikasikan sentimen (Positif, Negatif, dan Netral) dari 50.954 komentar pada video pembelajaran bahasa Inggris di *YouTube* menggunakan algoritma *Naïve Bayes Classifier* (NBC).

Metodologi penelitian ini melibatkan serangkaian tahap text mining yang komprehensif, mulai dari pengumpulan data melalui *YouTube Data API v3*, *preprocessing* (termasuk *cleaning*, *case*

folding, stopword removal, dan stemming), *labelling* otomatis menggunakan VADER *Sentiment Analyzer*, hingga ekstraksi fitur menggunakan *TF-IDF*.

Kesimpulan Utama:

1. Efektivitas Model: Algoritma NBC terbukti efektif untuk klasifikasi sentimen pada data komentar ini dengan mencapai Akurasi Keseluruhan sebesar 84%. Kinerja yang solid ini didukung oleh proses *preprocessing* dan ekstraksi fitur *TF-IDF* yang optimal.
2. Dominasi Sentimen Netral: Hasil pelabelan menunjukkan adanya kelas yang tidak seimbang (*class imbalance*) yang signifikan, yang mana sentimen Netral mendominasi (56.34%), diikuti Positif (32.77%), dan Negatif (10.89%).
3. Tantangan Klasifikasi: Meskipun akurasi keseluruhan tinggi, kinerja model pada kelas minoritas, yaitu sentimen Negatif, adalah yang terlemah (*Recall* hanya 0.41). Hal ini disebabkan oleh bias model ke kelas mayoritas dan kesalahan klasifikasi di mana banyak komentar yang sesungguhnya Negatif justru salah diprediksi sebagai Positif (313 kasus).

Secara keseluruhan, *Naïve Bayes Classifier* adalah solusi yang valid untuk analisis sentimen ini. Namun, disarankan bagi penelitian selanjutnya untuk mempertimbangkan teknik penanganan *class imbalance* agar kemampuan model dalam mendeteksi dan mengklasifikasikan sentimen Negatif dapat ditingkatkan secara lebih akurat. Hal ini penting bagi pembuat konten untuk mendapatkan umpan balik kritis yang lebih andal guna memaksimalkan kualitas video pembelajaran mereka.

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi dan ucapan terima kasih setulus-tulusnya kepada seluruh pihak yang terlibat. Dukungan, arahan, serta curahan waktu, tenaga, dan pikiran yang telah diberikan

merupakan bekal tak ternilai dalam menyelesaikan studi dan penelitian ini.

DAFTAR PUSTAKA

- Al, H., Harpizon, R., Kurniawan, R., & Iskandar, I. (2022). *Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes*. 5(1), 131–140.
- Anam, K. (2025). *Comparison of Sentiment Labeling Using Textblob, Vader, and Flair in Public Opinion Analysis Post-2024 Presidential Inauguration with IndoBERT*. 6(2), 803–818.
- Aqsalia, Z. Y., Sari, S., & Umi, N. (2024). *Optimasi Algoritma Naive Bayes Untuk Menganalisis Sentimen Pada Konten Pemandangan Ibu Kota di Youtube*. 5(2), 68–83.
- Arrahman Firsttama, R., Anjani Arifiyanti, A., & Satria Yudha Kartika, D. (2024). *Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes*. 6(2), 282–285.
- Fatah, Z. (2025). *TIK dan Masyarakat* (T. P. P. Media (ed.); first). PT Penamuda Media.
- Febrialdi Ansyah, M., Ghofur, A., & Fakhri Lidimillah, L. (2024). *G-Tech : Jurnal Teknologi Terapan*. 8(2), 847–856.
- Huwaida, S. F., Kusumawati, R., & Isnaini, B. (2024). *Analisis sentimen komentar youtube terhadap pemindahan ibu kota negara menggunakan metode Naïve Bayes*. 6(1), 26–39. <https://doi.org/10.37905/jji.v6i1.24718>
- Jamil, M., Hadiyanto, H., & Sanjaya, R. (2024). *Ingénierie des Systèmes d'Information Sentiment Analysis : Classifying Public Comments on YouTube in Disaster Management Simulation in Indonesia Using Naïve Bayes and Support Vector Machine*. 29(2), 437–446.
- Jelita, H. P., & Saad, M. I. (2025). *Penerapan Algoritma Naïve Bayes Dalam Analisis Sentimen Masyarakat*

- Terhadap STMIK Widya Cipta Dharma.* 6(2), 148–160. <https://doi.org/10.47065/bit.v5i2.2029>
- Khanafiyah, Y., Informatika, T., Group, S., Informatika, T., Group, S., Sari, D. K., Informatika, T., Group, S., Bayes, N., & Vote, D. (2025). *Analisis Sentimen Komentar Youtube Kanal Dirty Vote Menggunakan Metode Naive Bayes Classifier.* 12(4), 6723–6728.
- Mareta, A., Subroto, D. E., Aulia, L., & Nuryanah, S. (2025). *Peran Media Sosial Youtube sebagai Media Edukasi dalam Pendidikan Generasi Z.*
- Misrun, C. A., Haerani, E., Fikry, M., & Budianita, E. (2024). *Jurnal Computer Science and Information Technology (CoSciTech) naive bayes classifier method.* 4(1), 207–215.
- Morgan, P., Show, B., & Media, N. (2022). *Analisis Sentimen Pengguna Youtube Terhadap Tayangan # Matanajwamentiterawan Dengan Metode Naive Bayes Sentiment Analysis Of User Comments On Youtube Video # Matanajwamentiterawan With Naive Bayes Classifier Method.* 5(1), 1–6. <https://doi.org/10.33387/jiko>
- Nur Hidayah, W. (2021). *Jurnal Pendidikan Indonesia (Japendi) Pendidikan Konsep Solusi Terhadap Problem Keterampilan (Skills) Berbahasa Inggris Di Lembaga Pendidikan Indonesia Widha Nur Hidayah * Universitas Islam Negeri Sunan Kalijaga Yogyakarta , Indonesia Email : widha.hi.* 2(10), 1824–1834.
- Polus, M. E. (2021). *Development For Performance Of Porter Stemmer Algorithm.* 2, 6–13. <https://doi.org/10.15587/1729-4061.2021.225362>
- Putri, A. A., Bautista, C., Irsyad, H., Rahman, A., Multi, U., Palembang, D., Selatan, S., Multi, U., Palembang, D., Selatan, S., Multi, U., Palembang, D., Selatan, S., Multi, U., Palembang, D., & Selatan, S. (2025). *Opini Publik Terhadap Isu Keaslian Ijazah Pada Platform Youtube Dengan Naive Bayes , Knn , Dan.* 11(2), 231–242.
- Rahman, A., Rahmat, F., Fariqi, M. Y., & Adi, S. (2020). *Metode Naive Bayes untuk Menganalisis Akurasi Sentimen Komentar di Youtube.* 14(1), 31–34.
- Taufiqurahman, M. D., Anraeni, S., & Darwis, H. (2024). *Analisis Sentimen Komentar Konten Kreator Gaming Menggunakan Metode Naive Bayes dan KNN.* 1(4), 317–327.
- Yasir, M., & Suraji, R. (2023). *Perbandingan Metode Klasifikasi Naive Bayes , Decision Tree , Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 Pada Media Sosial Youtube Muhammad Yasir ¹ , Robertus Suraji² Perbandingan Metode Klasifikasi Naive Bayes , Decision Tree , Ran.*