

**ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP
APLIKASI TEMU DAN REAKSI PENOLAKAN
PEMERINTAH MENGGUNAKAN ALGORITMA NAÏVE
BAYES**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Informatika



disusun oleh
KAVI RAJA
21.11.3904

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP
APLIKASI TEMU DAN REAKSI PENOLAKAN
PEMERINTAH MENGGUNAKAN ALGORITMA NAÏVE
BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Informatika



disusun oleh
KAVI RAJA
21.11.3904

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP
APLIKASI TEMU DAN REAKSI PENOLAKAN PEMERINTAH
MENGUNAKAN ALGORITMA NAÏVE BAYES**

yang disusun dan diajukan oleh

Kavi Raja

21.IT.3904

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Juni 2025

Dosen Pembimbing,



Dwi Nurani, S.Kom., M.Kom.

NIK. 190302236

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP APLIKASI
TEMU DAN REAKSI PENOLAKAN PEMERINTAH MENGGUNAKAN
ALGORITMA NAÏVE BAYES**

yang disusun dan diajukan oleh

Kavi Raja

21.11.3904

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Juni 2025

Susunan Dewan Penguji

Nama Penguji

Majid Rahardi, S.Kom., M.Eng.

NIK. 190302393

Irma Rofni Wulandari, S.Pd., M.Eng.

NIK. 190302329

Dwi Nurani, S.Kom., M.Kom.

NIK. 190302236

Tanda Tangan



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Juni 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Kavi Raja
NIM : 21.11.3904

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Pengguna Youtube Terhadap Aplikasi Temu dan Reaksi
Penolakan Pemerintah Menggunakan Naïve Bayes**

Dosen Pembimbing : Dwi Nurani, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Juni 2025

Yang Menyatakan,



Kavi Raja

HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati, karya ini saya persembahkan kepada:

1. Tuhan Yang Maha Esa, sumber segala kekuatan dan petunjuk dalam setiap Langkah kehidupan saya.
2. Orang Tua saya tercinta, yang selalu menjadi sumber semangat, doa, dan pengorbanan tanpa batas.
3. Keluarga, pasangan, sahabat, dan teman seperjuangan, yang telah menjadi system pendukung terbaik dalam perjalanan kehidupan saya.
4. Dosen pembimbing dan penguji, atas bimbingan, ilmu, dan waktunya yang sangat berarti.
5. Diri saya sendiri, atas keberanian untuk terus berjuang meski suka menunda namun tidak menyerah di tengah jalan.

Semoga karya ini menjadi wujud rasa syukur dan awal dari Langkah yang lebih besar di masa depan.

KATA PENGANTAR

Puji syukur saya panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas berkat, rahmat, dan karunia-Nya, saya dapat menyelesaikan skripsi yang berjudul: “Analisis Sentimen Pengguna Youtube Terhadap Aplikasi Temu dan Reaksi Penolakan Pemerintah Menggunakan Algoritma Naïve Bayes”

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Penyusunan skripsi ini tentunya tidak terlepas dari bantuan, dukungan, serta bimbingan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, saya menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Dosen pembimbing yang telah membimbing dengan sabar serta memberikan ilmu yang sangat berarti selama proses penulisan skripsi ini.
2. Dosen penguji yang telah memberikan kritik dan masukan yang membangun demi penyempurnaan skripsi ini.
3. Kedua orang tua saya, atas cinta, doa, dukungan moral dan material yang tidak pernah henti diberikan sepanjang perjalanan ini.
4. Keluarga, pasangan, sahabat, dan rekan seperjuangan, atas semangat, canda tawa, serta motivasi yang turut menjadi bagian penting dalam proses ini.

Saya menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, saya sangat terbuka terhadap kritik dan saran yang membangun untuk perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat dan menjadi referensi bagi penelitian selanjutnya dalam bidang analisis sentimen dan pemrosesan bahasa alami.

Yogyakarta, 19 Juni 2025

Penulis

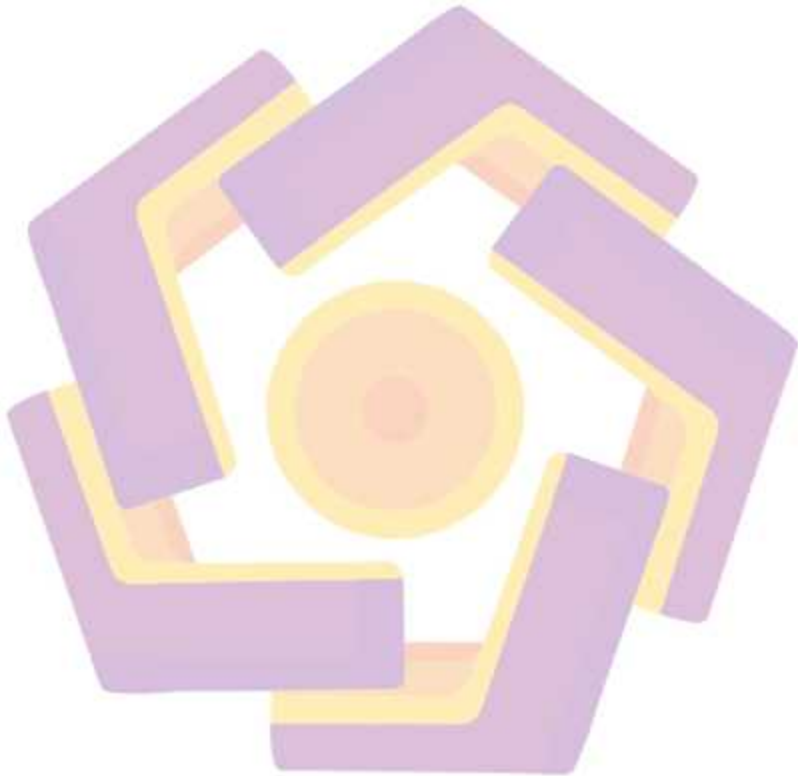
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	Error! Bookmark not defined.
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	14

2.2.1	Sentimen Analisis	14
2.2.2	Media Sosial Youtube.....	14
2.2.3	Aplikasi Temu dan Reaksi Penolakan Pemerintah	15
2.2.4	Machine Learning	15
2.2.5	Python	17
2.2.6	Naïve Bayes	18
2.2.7	TF-IDF	20
2.2.8	Imbalanced Dataset	21
2.2.9	Evaluasi Model	22
BAB III METODE PENELITIAN		25
3.1	Alur Penelitian	25
3.2	Alat dan Bahan.....	30
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Pengumpulan Data	34
4.2	Exploratory Data Analysis.....	35
4.3	Preprocessing	44
4.4	TF-IDF	52
4.5	SMOTE	54
4.6	Training Model dan Evaluasi Model	55
BAB V PENUTUP		60
5.1	Kesimpulan	60
5.2	Saran	61
REFERENSI		63
LAMPIRAN.....		66

DAFTAR TABEL

Tabel 2. 1 Tabel Keaslian	10
Tabel 3. 1 Daftar video Youtube sebagai sumber data komentar	30



DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	25
Gambar 4. 1 Data yang berhasil dikumpulkan.....	34
Gambar 4. 2 Informasi dan kelengkapan kolom.....	35
Gambar 4. 3 Pengecekan duplikasi pada data.....	36
Gambar 4. 4 Data setelah seleksi kolom.....	36
Gambar 4. 5 Hasil pemeriksaan data kosong.....	37
Gambar 4. 6 Hasil penanganan data kosong.....	37
Gambar 4. 7 Proses pengubahan nama kolom.....	37
Gambar 4. 8 Distribui jumlah komentar berdasarkan label.....	38
Gambar 4. 9 Distribusi dalam bentuk presentase.....	38
Gambar 4. 10 Visualisasi data distribusi komentar.....	38
Gambar 4. 11 Histogram distribusi jumlah karakter pada komentar.....	39
Gambar 4. 12 Histogram distribusi jumlah kata per komentar.....	40
Gambar 4. 13 Analisis rata-rata panjang karakter dan jumlah kata perlabel.....	41
Gambar 4. 14 Analisis berdasarkan kategori.....	41
Gambar 4. 15 Visualisasi analisis berdasarkan kateogori.....	42
Gambar 4. 16 Wordcloud sentimen negatif.....	43
Gambar 4. 17 WordCloud sentimen netral.....	43
Gambar 4. 18 Wordcloud sentimen positif.....	43
Gambar 4. 19 Hasil proses case folding.....	44
Gambar 4. 20 Hasil proses cleaning teks.....	45
Gambar 4. 21 Topword Negatif.....	45
Gambar 4. 22 Topword positif.....	46
Gambar 4. 23 Topword netral.....	46
Gambar 4. 24 Hasil normalisasi kata.....	47
Gambar 4. 25 Hasil tokenisasi.....	47
Gambar 4. 26 Daftar kata yang dipertahankan dan dihilangkan.....	49

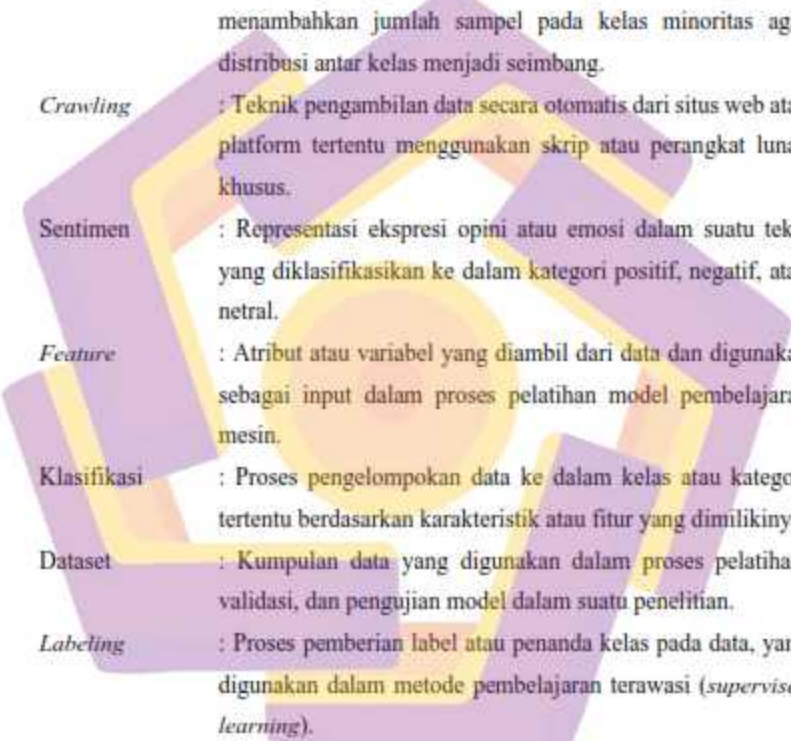
Gambar 4. 27 Hasil <i>stopword removal</i>	49
Gambar 4. 28 <i>Wordcloud after stopwords removal</i>	50
Gambar 4. 29 Kata yang dipertahankan pada proses <i>stemming</i>	51
Gambar 4. 30 Hasil proses <i>stemming</i>	51
Gambar 4. 31 Struktur dataset setelah <i>preprocessing</i>	52
Gambar 4. 32 Distribusi data setelah tahap <i>preprocessing</i>	52
Gambar 4. 33 Hasil pengestrakan TF-IDF	53
Gambar 4. 34 Hasil transformasi TF-IDF	53
Gambar 4. 35 Hasil hitung tingkat kekosongan.....	54
Gambar 4. 36 Hasil <i>split</i> Data	54
Gambar 4. 37 Hasil penerapan SMOTE	54
Gambar 4. 38 Hasil evaluasi model MNB menggunakan data asli.....	55
Gambar 4. 39 Hasil evaluasi model CNB menggunakan data asli	56
Gambar 4. 40 Hasil evaluasi model MNB menggunakan data SMOTE	56
Gambar 4. 41 Hasil evaluasi model CNB menggunakan data SMOTE	57
Gambar 4. 42 Menentukan alpha terbaik pada model menggunakan data asli.....	57
Gambar 4. 43 Menentukan alpha terbaik pada model menggunakan data SMOTE	57
Gambar 4. 44 Hasil evaluasi tuning MNB model menggunakan data asli	58
Gambar 4. 45 Hasil evaluasi tuning model CNB menggunakan data asli	58
Gambar 4. 46 Hasil evaluasi tuning model MNB menggunakan data SMOTE	59
Gambar 4. 47 Hasil evaluasi tuning model CNB menggunakan data SMOTE	59

DAFTAR LAMBANG DAN SINGKATAN



API	: <i>Application Programming Interface</i>
CSV	: <i>Comma-Separated Values</i>
D2C	: <i>Direct to Consumer</i>
C2M	: <i>Consumer to Manufacturer</i>
EDA	: <i>Exploratory Data Analysis</i>
MNB	: <i>Multinomial Naïve Bayes</i>
CNB	: <i>Complement Naïve Bayes</i>
NLP	: <i>Natural Language Processing</i>
SMOTE	: <i>Synthetic Minority Over-sampling Technique</i>
TF	: <i>Term Frequency</i>
IDF	: <i>Inverse Document Frequency</i>
TF-IDF	: <i>Term Frequency – Inverse Document Frequency</i>
FN	: <i>False Negative</i>
FP	: <i>False Positive</i>
TN	: <i>True Negative</i>
TP	: <i>True Positive</i>
UMKM	: <i>Usaha Mikro, Kecil, dan Menengah</i>
VSCoDe	: <i>Visual Studio Code</i>

DAFTAR ISTILAH



<i>Preprocessing</i>	: Proses awal dalam pengolahan data teks yang bertujuan untuk membersihkan, menyederhanakan, dan menyiapkan data agar dapat dianalisis secara optimal.
<i>Oversampling</i>	: Teknik penanganan ketidakseimbangan data dengan cara menambahkan jumlah sampel pada kelas minoritas agar distribusi antar kelas menjadi seimbang.
<i>Crawling</i>	: Teknik pengambilan data secara otomatis dari situs web atau platform tertentu menggunakan skrip atau perangkat lunak khusus.
<i>Sentimen</i>	: Representasi ekspresi opini atau emosi dalam suatu teks, yang diklasifikasikan ke dalam kategori positif, negatif, atau netral.
<i>Feature</i>	: Atribut atau variabel yang diambil dari data dan digunakan sebagai input dalam proses pelatihan model pembelajaran mesin.
<i>Klasifikasi</i>	: Proses pengelompokan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik atau fitur yang dimilikinya.
<i>Dataset</i>	: Kumpulan data yang digunakan dalam proses pelatihan, validasi, dan pengujian model dalam suatu penelitian.
<i>Labeling</i>	: Proses pemberian label atau penanda kelas pada data, yang digunakan dalam metode pembelajaran terawasi (<i>supervised learning</i>).
<i>Overfitting</i>	: Kondisi di mana model terlalu menyesuaikan terhadap data pelatihan sehingga performanya menurun saat diuji pada data baru.
<i>Split Data</i>	: Prosedur pembagian dataset menjadi bagian-bagian tertentu, seperti data pelatihan dan data pengujian, untuk keperluan evaluasi model.

- Tuning* : Proses penyetelan parameter pada model pembelajaran mesin guna memperoleh performa terbaik berdasarkan metrik evaluasi tertentu.
- Noise* : Elemen-elemen pada data yang tidak relevan atau mengganggu proses analisis, seperti karakter tidak penting, kesalahan penulisan, atau simbol asing.
- Slang* : Kata-kata tidak baku atau informal yang umum digunakan dalam komunikasi sehari-hari dan sering kali memerlukan normalisasi dalam analisis teks.



INTISARI

Pertumbuhan *e-commerce* di Indonesia telah memicu munculnya berbagai platform baru, salah satunya adalah aplikasi Temu yang menuai kontroversi karena dianggap mengancam pelaku UMKM lokal. Penolakan pemerintah terhadap Temu memunculkan reaksi dan perdebatan di media sosial, khususnya Youtube. Penelitian ini bertujuan untuk menganalisis sentimen pengguna Youtube terhadap aplikasi Temu dan reaksi penolakan pemerintah dengan menerapkan algoritma Naïve Bayes. Data dikumpulkan melalui *crawling* komentar pada tujuh video Youtube yang relevan, menghasilkan 6.090 komentar yang kemudian diberi label sentimen secara manual ke dalam tiga kategori yaitu positif, negatif, dan netral dengan mayoritas data yang condong ke label negatif yang berjumlah 60,14% dari total data yang ada. Tahapan penelitian meliputi eksplorasi data, *preprocessing* yang mencakup *case folding*, *cleaning text*, normalisasi, tokenisasi, *stopword removal*, *stemming*. Selanjutnya ekstraksi fitur menggunakan TF-IDF, penyeimbangan data dengan teknik SMOTE, pelatihan model Multinomial dan Complement Naïve Bayes, serta evaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa mayoritas sentimen pengguna bersifat negatif (60,14%), dengan akurasi model terbaik mencapai lebih dari 74% pada model MNB setelah penerapan SMOTE dan *tuning hyperparameter*. Penelitian ini memberikan kontribusi dalam pengembangan metode analisis sentimen berbasis *machine learning* dan dapat dimanfaatkan oleh pemerintah, pelaku bisnis, serta peneliti untuk memahami persepsi publik terhadap isu *e-commerce* dan kebijakan digital di Indonesia.

Kata kunci: analisis sentimen, Naïve Bayes, Youtube, SMOTE.

ABSTRACT

The rapid growth of e-commerce in Indonesia has triggered the emergence of various new platforms, one of which is the Temu application, which has sparked controversy due to concerns that it threatens local MSMEs. The Indonesian government's rejection of Temu has generated reactions and debates on social media, particularly on YouTube. This study aims to analyze YouTube users' sentiment toward the Temu application and the government's rejection response by applying the Naïve Bayes algorithm. The data were collected by crawling comments from seven relevant YouTube videos, resulting in 6,090 comments that were manually labeled into three sentiment categories: positive, negative, and neutral. The majority of the data leaned toward the negative label, accounting for 60.14% of the total. The research stages include data exploration, preprocessing—which covers case folding, text cleaning, normalization, tokenization, stopword removal, and stemming—followed by feature extraction using TF-IDF, data balancing using the SMOTE technique, model training using Multinomial and Complement Naïve Bayes, and evaluation using accuracy, precision, recall, and F1-score metrics. The results show that the majority of user sentiment is negative (60.14%), with the best model accuracy exceeding 74% on the Multinomial Naïve Bayes model after applying SMOTE and hyperparameter tuning. This study contributes to the development of machine learning-based sentiment analysis methods and can be utilized by the government, business actors, and researchers to better understand public perception of e-commerce issues and digital policy in Indonesia.

Keywords: *sentiment analysis, Naïve Bayes, Youtube, SMOTE.*