

ANALISIS HUBUNGAN PENGGUNAAN BAHASA NONBAKU DAN SENTIMEN KOMENTAR YOUTUBE PADA KASUS KECURANGAN SAAT UTBK

Amalia Azzahra¹, Salwa Risda Pratiwi², Wulan Fransiska Lubis³, Floren Green Chia⁴, Veronika Valentin S. Garingging⁵, Elvi Simanjuntak⁶

¹⁻⁶Program Studi Statistika, Universitas Negeri Medan

¹amaliaazzahra.4242260001@mhs.unimed.ac.id,

²salwarisda.4241260005@mhs.unimed.ac.id,

³wulan.4241260014@mhs.unimed.ac.id, ⁴floren.4243260010@mhs.unimed.ac.id,

⁵vera.4243260008@mhs.unimed.ac.id,

⁶elvijuntak.4241260007@mhs.unimed.ac.id

ABSTRACT

This study examines the use of non-standard language in YouTube comment sections regarding cheating during Computer-Based Written Exams and its relationship to public sentiment in the comments. The study aims to identify the forms of non-standard language use and the relationship between the number of non-standard words and comment sentiment. The research method used is quantitative descriptive with a sociolinguistic approach. The research data consists of 390 YouTube comments obtained through automated data collection techniques. Analysis was conducted through the stages of data preprocessing, identification of non-standard words, sentiment analysis, calculating the frequency of non-standard word occurrences, creating a cross-tabulation table, and a chi-square test. The results of the study indicate that negative sentiment has the highest usage of non-standard words, with a chi-square test result of 0.000024, indicating a significant relationship between the use of non-standard words and public sentiment in social media comments.

Keywords: Sentiment Analysis, Non-Standard Language, UTBK, YouTube

ABSTRAK

Penelitian ini membahas penggunaan bahasa nonbaku pada kolom komentar YouTube terkait tindakan kecurangan selama Ujian Tulis Berbasis Komputer serta hubungannya dengan sentimen komentar masyarakat. Penelitian bertujuan untuk mengetahui bentuk penggunaan bahasa nonbaku dan hubungan antara jumlah kata tidak baku dengan sentimen komentar. Metode penelitian yang digunakan adalah deskriptif kuantitatif dengan pendekatan sosiolinguistik. Data penelitian berupa 390 komentar YouTube yang diperoleh melalui teknik pengambilan data secara otomatis. Analisis dilakukan melalui tahap preprocessing data, identifikasi kata tidak baku, analisis sentimen, menghitung frekuensi kemunculan data tidak baku, membuat tabel silang, dan uji chi-square. Hasil penelitian menunjukkan bahwa sentimen negatif memiliki penggunaan kata tidak baku paling tinggi dengan hasil uji chi-square 0,000024 menunjukkan terdapat hubungan signifikan antara

penggunaan kata tidak baku dan sentimen komentar masyarakat pada media sosial.

Kata Kunci: Analisis Sentimen, Bahasa Nonbaku, UTBK, YouTube

A. Pendahuluan

Bahasa adalah sarana komunikasi dan interaksi lisan yang khusus dimiliki oleh manusia. Bahasa dapat dijelaskan dari sudut pandang internal dan eksternal. Analisis internal berarti mengkaji hanya pada struktur dalam bahasa itu sendiri, termasuk aspek fonologi, morfologi, sintaksis, hingga wacana. Di sisi lain, analisis eksternal merupakan kajian terhadap aspek atau faktor-faktor di luar bahasa, namun masih berkaitan dengan cara penggunaan bahasa oleh para penuturnya dalam kelompok sosial masyarakat (Kaharuddin et al., 2022).

Di tengah perubahan yang tidak bisa dihindari ini, bahasa tidak hanya dipengaruhi oleh aspek-aspek linguistik, namun juga oleh beragam faktor sosial yang lebih rumit. Salah satu hal yang penting untuk dipahami dalam bidang linguistik adalah keberagaman bahasa. Dalam bidang sosiolinguistik, keberagaman bahasa adalah fokus utama dari penelitian. Sosiolinguistik mempelajari bagaimana bahasa digunakan di berbagai strata masyarakat, serta

bagaimana keberagaman bahasa yang ada mencerminkan struktur sosial, identitas individu dan kelompok, serta dinamika budaya di masyarakat tersebut. Keberagaman bahasa mencakup banyak aspek, seperti dialek, register, variasi bahasa, dan perubahan bahasa yang terjadi dalam masyarakat (Sulistyawati et al., 2025).

Menurut Badan Pengembangan dan Pembinaan Bahasa (2021), terdapat tren penurunan penggunaan bahasa Indonesia baku oleh remaja, terutama dalam komunikasi dan tulisan akademik. Penggunaan bahasa non-baku yang berlebihan dapat mengancam keberlangsungan bahasa Indonesia sebagai bahasa yang berfungsi untuk mempersatukan dan membentuk identitas nasional. Rasa bangga dan kepedulian terhadap bahasa Indonesia dapat berkurang jika generasi muda lebih suka berbicara dalam bahasa asing atau bahasa gaul (Sebayang et al., 2024).

Penggunaan bahasa informal dan non-baku di media sosial sering menyebabkan terjadinya tentang

apakah fenomena ini merupakan inovasi linguistik atau ancaman bagi keberlangsungan bahasa Indonesia sebagai bahasa nasional (Ulya, 2024). Dalam penggunaannya, bahasa Indonesia memiliki dua variasi, yaitu bahasa baku dan nonbaku. Bahasa baku adalah variasi bahasa yang cara berbicara dan menulisnya mengikuti aturan yang telah ditetapkan. Aturan ini bisa berupa panduan ejaan, tata bahasa resmi, dan kamus umum. Sebaliknya, bahasa non-baku adalah variasi bahasa yang cara penyebutan atau penulisannya tidak sesuai dengan aturan tersebut. Penggunaan variasi bahasa baku dan non-baku tergantung pada situasi dan konteks penggunaan. Variasi bahasa baku biasanya digunakan dalam konteks resmi, seperti dalam seminar, pidato, pertemuan ilmiah, dan sebagainya. Sementara itu, variasi bahasa non-baku yang umumnya dipakai dalam komunikasi sehari-hari yang tidak bersifat formal (Waridah, 2008).

Kemajuan teknologi digital, terutama platform media sosial yang fokus pada video seperti YouTube, telah membuka ruang baru untuk penggunaan bahasa dalam komunikasi publik. YouTube tidak hanya berfungsi sebagai platform

hiburan, tetapi juga sebagai alat pendidikan dan interaksi sosial yang memungkinkan terjadinya komunikasi dua arah antara pembuat konten dan penonton. Dalam konteks ini, penggunaan bahasa sering kali mencerminkan keragaman sosial dan situasi, menjadikannya menarik untuk dijelaskan dari sudut pandang sosiolinguistik (Sianipar et al., 2026).

Platform seperti Youtube tidak hanya berfungsi untuk membagikan video tetapi juga menjadi wadah diskusi publik melalui fitur komentar. Fenomena penggunaan bahasa gaul dalam kolom komentar Youtube sangat menarik untuk diteliti karena kolom komentar berfungsi sebagai ruang publik yang menjadi lokasi berkumpulnya beragam individu dari berbagai latar belakang sosial, usia, dan pendidikan yang berbeda. Hal ini membuat kolom komentar menjadi sarana interaksi sosial yang kaya akan variasi bahasa, termasuk bahasa gaul yang mencerminkan nilai budaya dan sosial yang tengah berkembang dalam masyarakat (Nurmainia et al., 2025)

Salah satu topik yang banyak dibahas masyarakat baru-baru ini adalah pelaksanaan Ujian Tulis Berbasis Komputer (UTBK). Menurut

Hidayat R. A., et al (dalam Syalsabella & Prapanca, 2024) proses pemilihan penerimaan mahasiswa ke universitas saat ini menggunakan metode yang sepenuhnya digital, sejalan dengan pesatnya kemajuan di bidang digital saat ini. Salah satunya adalah Ujian Tulis Berbasis Komputer (UTBK), yang merupakan tes yang dilaksanakan menggunakan komputer, dan hasilnya digunakan untuk proses seleksi masuk ke universitas negeri. Ujian ini telah diperkenalkan sejak tahun 2018 dan diorganisir oleh LTMPT (Lembaga Tes Masuk Perguruan Tinggi).

Dalam pelaksanaannya, UTBK juga tidak terlepas dari berbagai permasalahan seperti kecurangan yang dilakukan oleh peserta. Kecurangan adalah tindakan penipuan yang dilakukan dengan sengaja yang menyebabkan kerugian tanpa disadari oleh pihak yang dirugikan dan memberikan manfaat bagi pelaku kecurangan. Biasanya, kecurangan terjadi akibat adanya tekanan untuk melakukan penyimpangan atau dorongan untuk memanfaatkan peluang yang ada serta adanya justifikasi yang diterima secara luas terhadap tindakan tersebut (Natasia et al., 2021). Seperti

yang telah kita pahami, tujuan diadakannya ujian seleksi perguruan tinggi adalah untuk menilai kemampuan yang dimiliki oleh individu yang ingin menjadi mahasiswa baru agar dapat diterima di program studi yang sesuai dengan kapasitas mereka. Oleh sebab itu, tindakan menggunakan joki dalam tes tersebut adalah sesuatu yang salah. Hal ini disebabkan karena terdapat elemen penipuan di mana peserta ujian masuk perguruan tinggi melakukan tindakan pemalsuan dalam prosesnya (Dzaky, 2023).

Beberapa penelitian terdahulu memiliki keterkaitan dengan kajian ini. Penelitian pertama oleh (Rahmawati & Aini, 2025) yang berjudul “Analisis Sentimen Komentar YouTube pada Program *Clash of Champions* Ruangguru Menggunakan Deep Learning Berbasis LSTM” menunjukkan bahwa komentar di media sosial mengandung berbagai bentuk bahasa informal yang dapat dianalisis menggunakan pendekatan *Natural Language Processing* (NLP) untuk mengetahui kecenderungan sentimen publik. Selanjutnya, pada penelitian kedua oleh (Nurmainia et al., 2025) yang berjudul “Penggunaan Variasi Bahasa “Slang” pada Kolom

Komentar di Platform Media Sosial Youtube @Deddy Corbuzier & @Denny Sumargo: Kajian Sociolinguistik, mengungkapkan bahwa bahasa nonbaku muncul sebagai bentuk ekspresi yang dipengaruhi oleh faktor kepraktisan, kedekatan sosial, dan kebiasaan pengguna media sosial

Berdasarkan penelitian-penelitian tersebut, dapat diketahui bahwa penggunaan bahasa nonbaku dan analisis sentimen di media sosial telah banyak dikaji, namun penelitian mengenai fenomena joki UTBK di YouTube masih terbatas. Oleh karena itu, novelty penelitian ini terletak pada analisis transformasi bahasa nonbaku menjadi bahasa baku pada kolom komentar YouTube yang berkaitan dengan kasus joki UTBK, serta pengelompokan sentimen masyarakat ke dalam kategori positif, netral, dan negatif. Tujuan dari penelitian ini adalah untuk mengetahui apakah terdapat hubungan yang signifikan antara jumlah kata tidak baku dan sentimen komentar.

B. Metode Penelitian

Metode penelitian yang digunakan adalah pendekatan deskriptif kuantitatif dengan kajian

sociolinguistik. Pendekatan kuantitatif digunakan untuk memberikan gambaran yang jelas dan mendalam tentang data yang telah dikumpulkan sehingga lebih mudah untuk memahami dan membuat keputusan berdasarkan data ini (Sudirman et al., 2020). Objek dari penelitian ini adalah komentar YouTube yang mengandung bahasa nonbaku terkait kasus joki UTBK, dengan teknik pengumpulan data metode scrapping. Scraping data berarti mengambil data otomatis dari berbagai situs web dan mengubahnya menjadi format yang siap untuk digunakan dalam analisis data (Urva et al., 2023).

Langkah Analisis Data

Pada tahap analisis data, penelitian dilakukan dengan mendeteksi kata tidak baku pada setiap komentar YouTube yang telah melalui proses preprocessing. Berikut adalah langkah-langkah dalam analisis data:

1. Preprocessing Data

Preprocessing adalah langkah pertama dalam pengelompokan teks yang bertujuan untuk menyiapkan data teks sebelum diolah lebih jauh (Aufar et al., 2023). Menurut (Saputro et al., 2025), tahapan preprocessing

data terdiri dari beberapa langkah, tiga diantaranya meliputi (case folding, cleaning/pembersihan karakter khusus, dan tokenizing):

- a. *Case Folding*, adalah langkah untuk mengubah semua huruf dalam sebuah komentar menjadi huruf kecil. Tujuannya adalah untuk menyamakan kata-kata yang sama meski terdapat perbedaan dalam penggunaan huruf besar.
- b. *Cleaning*, adalah menghapus bagian-bagian yang tidak berkaitan dengan teks, seperti URL, sebutan, tagar, emoji, dan spasi ganda. Proses ini menjamin bahwa hanya konten utama yang dievaluasi sehingga hasil pengkategorian sentimen menjadi lebih tepat.
- c. *Tokenisasi* adalah langkah berikutnya untuk memisahkan komentar menjadi bagian-bagian kata (tokens) agar analisis dapat dilakukan secara individual. Prosedur ini diterapkan untuk mengenali kata-kata penting yang memiliki nilai sentimen dalam komentar.

2. Identifikasi dan Jumlah Kata Tidak Baku per Komentar

Pada tahap ini, komentar di YouTube dianalisis untuk mencari istilah yang tidak sesuai dengan aturan bahasa Indonesia baku. Proses identifikasi dilakukan dengan mencocokkan kata-kata dalam komentar dengan istilah resmi yang terdapat dalam KBBI. Setelah kata tidak baku ditemukan, jumlahnya dihitung pada setiap komentar. Tujuannya untuk mengetahui tingkat penggunaan bahasa nonbaku oleh masing-masing pengguna

3. Analisis Sentimen

Pelabelan sentimen merupakan metode untuk mengelompokkan tulisan berdasarkan sifat emosionalnya, seperti apakah itu positif, netral atau negatif. Tahapan klasifikasi emosi dilaksanakan dengan menilai nilai total, yang mencerminkan arah emosi keseluruhan dari tulisan tersebut (Saputra et al., 2025).

4. Menghitung Frekuensi

Kemunculan Kata Tidak Baku dan Membuat Tabel Silang

Tahap ini dilakukan untuk mengetahui kata tidak baku yang paling sering muncul dalam komentar. Hasil frekuensi dapat menunjukkan pola penggunaan bahasa pengguna media sosial. Data jumlah kata tidak baku dan kategori sentimen kemudian

disusun dalam bentuk tabel silang. Tabel ini digunakan untuk melihat hubungan antara penggunaan kata tidak baku dengan jenis sentimen komentar.

5. Melakukan Uji Chi-Square

Uji Chi-Square merupakan teknik statistik yang tidak bergantung pada asumsi tertentu, dan digunakan untuk menganalisis hubungan atau perbedaan antara dua atau lebih variabel yang bersifat kategorikal. Metode ini membandingkan data yang sebenarnya di lapangan dengan data yang diharapkan berdasarkan beberapa estimasi (Dalimunthe et al., 2025). Hipotesis dalam uji chi-square yaitu:

H_0 : tidak terdapat hubungan signifikan antara jumlah kata tidak baku dan sentimen komentar.

H_1 : terdapat hubungan signifikan antara kata tidak baku dan sentimen komentar.

Dasar dalam pengambilan keputusan uji chi-square yaitu tolak H_0 jika nilai signifikan $< 0,05$ artinya ada hubungan signifikan antara jumlah kata tidak baku dan sentimen komentar.

C. Hasil Penelitian dan Pembahasan

Bagian ini menyajikan hasil dan pembahasan dari penelitian tentang penggunaan bahasa nonbaku di kolom komentar YouTube yang berhubungan dengan kasus joki UTBK. Di bagian ini, disajikan rangkaian hasil dari pengolahan data awal, identifikasi, serta jumlah kata nonformal setiap komentar, analisis sentimen, dan pengujian hubungan antara penggunaan bahasa nonformal dengan emosi komentar menggunakan uji chi-square.

Data yang dikumpulkan berasal dari 9 video yang membahas tentang tindak kecurangan selama pelaksanaan UTBK. Proses pengambilan data dilakukan secara otomatis dengan menggunakan YouTube Comments Downloader, yang memungkinkan ekstraksi komentar dengan cara terorganisir. Jumlah total data yang berhasil dikumpulkan adalah 390 baris, di mana setiap komentar merepresentasikan satu komentar pengguna lengkap dengan informasi seperti nama pengguna, waktu publikasi dan jumlah suka. Contoh struktur data yang dikumpulkan disajikan pada Tabel 1.

Tabel 1. Struktur Data

Published	Author	Comment	Likes
12 days ago	@nurfah rudi3901	Mau jadi generasi apa,kalo kejujuran sudah hilang.	56
10 days ago	@hanna zzaaa	Mau peraturan pendaftaran UTBK sampe dirubah pun joki tetep aja masuk 😂😂 ampas emang kemendiktisain tek	3
9 days ago	@dzaky ammar379	bukannya itu oknum, banyak kok orang yang belajar dengan jujur. Lu malah bawa2 generasi	1
12 days ago	@lailatul muawanah	Kasian sama anak2 yg memang tes secara jujur mereka nunggu hasilnya dengan dag Dig dug semoga lulus dan bisa kuliah di salah satu fakultas faforitnya 🤔🤔	55
10 days ago	@Alima nShiddiqie	@SFwan. doakan aku juga bang, besok utbk	1

Preprocessing Data

Tahap preprocessing data dilakukan untuk membersihkan dan menyiapkan data komentar YouTube sebelum dilakukan analisis sentimen. Proses preprocessing meliputi *case folding*, *cleaning text*, dan *tokenizing*. Tahap ini bertujuan untuk menyeragamkan bentuk data teks agar lebih mudah dianalisis menggunakan

metode *Natural Language Processing* (NLP). Berikut contoh hasil preprocessing data komentar YouTube.

Tabel 2. Hasil Processing Data

<i>textDisplay</i>	<i>CaseFolding_text</i>	<i>clean_text</i>	<i>tokenized_text</i>
Mau jadi generasi apa,kalo kejujuran sudah hilang.	mau jadi generasi apa,kalo kejujuran sudah hilang.	mau jadi generas i apa kalo kejujuran sudah hilang	['mau', 'jadi', 'generasi', 'apakalo', 'kejujuran', 'sudah', 'hilang']
Mau peraturan pendaftaran UTBK sampe dirubah pun joki tetep aja masuk 😂😂 ampas emang kemendiktisaintek	mau peraturan pendaftaran utbk sampe dirubah pun joki tetep aja masuk 😂😂 ampas emang kemendiktisaintek	mau peraturan pendaftaran utbk sampe dirubah pun joki tetep aja masuk ampas emang kemendiktisaintek	['mau', 'peraturan', 'pendaftaran', 'utbk', 'sampe', 'dirubah', 'pun', 'joki', 'tetep', 'aja', 'masuk', 'ampas', 'emang', 'kemendiktisaintek']
bukannya itu oknum, banyak kok orang yang belajar dengan jujur. Lu malah bawa2 generasi	bukannya itu oknum, banyak kok orang yang belajar dengan jujur. lu malah bawa2 generasi	bukannya itu oknum banyak kok orang yang belajar dengan jujur lu malah bawa2 generasi	['bukannya', 'itu', 'oknum', 'banyak', 'kok', 'orang', 'yang', 'belajar', 'dengan', 'jujur', 'lu', 'malah', 'bawa2', 'generasi']
Kasian sama anak2 yg memang	kasian sama anak2 yg memang	kasian sama anak yg memang tes	['kasian', 'sama', 'anak', 'yg', 'memang']

tes	tes	secara	'tes',
secara	secara	jujur	'secara',
jujur	jujur	mereka	'jujur',
mereka	mereka	nunggu	'mereka',
nunggu	nunggu	hasilnya	'nunggu',
hasilnya	hasilnya	dengan	'hasilnya'
dengan	dengan	dag dig	,
dag Dig	dag dig	dug	'dengan',
dug	dug	semoga	'dag',
semoga	semoga	lulus	'dig',
lulus	lulus	dan bisa	'dug',
dan bisa	dan bisa	kuliah di	'semoga'
kuliah di	kuliah di	salah	, 'lulus',
salah	salah	satu	'dan',
satu	satu	fakultas	'bisa',
fakultas	fakultas	faforitny	'kuliah',
faforitny	faforitny	a	'di',
a 🙄🙄	a 🙄🙄		'salah',
			'satu',
			'fakultas',
			'faforitny
			a']
		doakan	['doakan'
@SFwa	@sfwan.	aku juga	, 'saya',
n.	doakan	bang	'juga',
doakan	aku juga	besok	'bang',
aku juga	bang,	utbk	'besok',
bang,	besok		'utbk']
besok	utbk		
utbk			

Identifikasi dan Jumlah Kata Tidak Baku per Komentar

Pada tahap ini, penggunaan kata tidak baku diidentifikasi pada komentar YouTube yang berkaitan dengan kecurangaan selama pelaksanaan UTBK. Komentar dengan jumlah kata tidak baku kurang dari satu dikategorikan sebagai "sedikit", dan komentar dengan jumlah kata tidak baku lebih dari satu dikategorikan sebagai "banyak". Pengelompokan ini dibuat untuk membuat lebih mudah untuk menganalisis hubungan antara

sentimen komentar dan penggunaan bahasa nonbaku.

Tabel 3. Jumlah Kata Tidak Baku per Komentar

<i>N</i> <i>o</i>	<i>Komentar</i>	<i>Jumla</i> <i>h</i> <i>Kata</i> <i>Tidak</i> <i>Baku</i>	<i>Kateg</i> <i>ori</i>
1	"Mau jadi generasi apa,kalo kejujuran sudah hilang.	1	Sedikit
2	"Mau peraturan pendaftaran UTBK sampe dirubah pun joki tetep aja masuk ampas emang kemendiktisaintek"	5	Banyak
3	"bukannya itu oknum, banyak kok orang yang belajar dengan jujur. Lu malah bawa2 generasi"	2	Banyak
4	"Kasian sama anak2 yg memang tes secara jujur mereka nunggu hasilnya dengan dag Dig dug semoga lulus dan bisa kuliah di salah satu fakultas faforitnya"	5	Banyak
5	"doakan saya juga bang, besok utbk"	1	Sedikit

Analisis Sentimen

Tahap selanjutnya adalah analisis sentimen. Analisis ini dilakukan untuk menentukan arah

emosional komentar masyarakat terhadap tindak kecurangan selama pelaksanaan UTBK. Proses ini dilakukan dengan mengelompokkan komentar ke dalam kategori sentimen positif, netral, dan negatif berdasarkan isi dan makna komentar pengguna. Penentuan label sentimen dilakukan berdasarkan nilai *confidence score* tertinggi yang dihasilkan model, sehingga setiap komentar akan diklasifikasikan ke dalam sentimen yang memiliki tingkat keyakinan prediksi paling tinggi.

Tabel 4. Hasil Analisis Sentimen

<i>N</i> <i>o</i>	<i>Komentar</i>	<i>Skor</i> <i>Sentimen</i>	<i>Label</i> <i>Sentimen</i>
1	"Mau jadi generasi apa,kalo kejujuran sudah hilang.	0.997493 982	Negatif
2	"Mau peraturan pendaftaran UTBK sampe dirubah pun joki tetep aja masuk ampas emang kemendiktisain tek"	0.996329 367	Negatif
3	"bukannya itu oknum, banyak kok orang yang belajar dengan jujur. Lu malah bawa2 generasi"	0.991201 52	Negatif
4	"Kasian sama anak2 yg memang tes secara jujur mereka nunggu	0.602046 788	Negatif

	hasilnya dengan dag Dig dug semoga lulus dan bisa kuliah di salah satu fakultas faforitnya"		
5	"doakan saya juga bang, besok utbk"	0.363760 442	Positif

Frekuensi Kata Tidak Baku dan Tabel Silang

Berdasarkan hasil analisis, diperoleh frekuensi penggunaan kata tidak baku pada masing–masing kategori sentimen sebagai berikut.

Tabel 5. Frekuensi Kata Tidak Baku

Sentimen	Total Kata Tidak Baku
Negatif	1.452
Netral	58
Positif	165

Sebelum dilakukan analisis tabulasi silang, jumlah kata tidak baku pada setiap komentar diklasifikasikan terlebih dahulu ke dalam dua kategori, yaitu sedikit dan banyak. Kategori sedikit diberikan pada komentar yang memiliki jumlah kata tidak baku ≤ 1 , sedangkan kategori banyak diberikan pada komentar yang memiliki jumlah kata tidak baku > 1 . Pengelompokan ini dilakukan untuk mempermudah analisis hubungan antara penggunaan kata tidak baku dengan kategori sentimen komentar. Hasil tabulasi silang antara kategori penggunaan kata tidak baku dan sentimen

komentar ditunjukkan pada tabel berikut.

Tabel 6. Hasil Tabulasi Silang

Kategori	Sentimen		
	Negatif	Netral	Positif
Banyak	244	13	24
Sedikit	79	21	9

Berdasarkan hasil frekuensi penggunaan kata tidak baku, sentimen negatif memiliki jumlah kata tidak baku tertinggi, yaitu sebanyak 1.452 kata, sedangkan sentimen netral memiliki jumlah paling rendah, yaitu 58 kata. Sentimen positif memiliki jumlah kata tidak baku sebanyak 165 kata.

Selanjutnya, hasil tabulasi silang menunjukkan bahwa kategori penggunaan kata tidak baku banyak didominasi oleh sentimen negatif sebanyak 244 komentar. Sementara itu, pada kategori sedikit, sentimen negatif juga memiliki jumlah tertinggi yaitu sebanyak 79 komentar. Hasil tersebut menunjukkan bahwa komentar dengan sentimen negatif cenderung menggunakan kata tidak baku dalam jumlah yang lebih banyak dibandingkan sentimen positif maupun netral. Hal ini mengindikasikan bahwa penggunaan bahasa tidak baku pada media sosial

sering digunakan untuk mengekspresikan emosi, kritik, dan ketidakpuasan terhadap suatu isu.

Uji Chi Square

Setelah dilakukan analisis tabulasi silang, selanjutnya melakukan uji Chi-Square untuk mengetahui apakah terdapat hubungan yang signifikan antara jumlah kata tidak baku dan sentimen komentar. Hasil uji Chi-Square menunjukkan nilai p-value sebesar 0,000024. Karena nilai p-value < 0,05, maka H_1 diterima. Hal ini menunjukkan bahwa terdapat hubungan yang signifikan antara jumlah kata tidak baku dan sentimen komentar. Dengan demikian, penggunaan kata tidak baku pada komentar memiliki hubungan dengan sentimen yang muncul, baik positif, netral, maupun negatif.

D. Kesimpulan

Penelitian ini menunjukkan bahwa penggunaan bahasa nonbaku dalam komentar YouTube mengenai kasus kecurangan UTBK memiliki hubungan yang signifikan dengan sentimen masyarakat. Melalui analisis pada 390 komentar, terlihat bahwa

sentimen negatif sangat mendominasi sebanyak 244 komentar dan memiliki frekuensi penggunaan kata tidak baku yang paling tinggi dibandingkan dengan sentimen positif dan netral. Hasil dari uji Chi-Square yang menunjukkan p-value 0,000024 menegaskan adanya hubungan yang signifikan antara frekuensi kata tidak baku dengan sentimen dalam komentar. Penelitian ini menyimpulkan bahwa bahasa yang nonbaku sering digunakan di media sosial sebagai sarana untuk mengekspresikan emosi, kritik, dan ketidakpuasan masyarakat tentang masalah yang sedang dibahas.

DAFTAR PUSTAKA

- Aufar, A. F., Rosid, M. A., Eviyanti, A., & Ika Ratna Indra Astutik. (2023). Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews. (*JICTE*) *Journal of Information and Computer Technology Education*, 7(2), 42–50.
<https://doi.org/10.21070/jicte.v7i2.1650>
- Dalimunthe, N. A., Nasution, A. R., & Amir, A. (2025). ANALISIS PENGGUNAAN UJI CHI-SQUARE PADA PENELITIAN PENDIDIKAN: STUDI LITERATUR. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, 10(04), 287–295.
- Dzaky, N. (2023). PENERAPAN SANKSI PIDANA PADA KASUS JOKI SELEKSI MASUK PERGURUAN TINGGI. *Jurnal Legisla*, 15, 14–24.
- Kaharuddin, Halijah, & B, E. (2022). Menganalisis Kemampuan Berbahasa Baku dan Berbahasa Tidak Baku bagi Siswa SMP Negeri 32 Bulukumba. *Elementary Journal*, 5(2), 64–74.
- Natasia, B., Aprilia, D., Oktaviyanti, D., Setiawan, D., Fadila, F. N., & Meikhati, E. (2021). ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TERJADINYA FRAUD DAALAM PELAPORAN KEUANGAN. *SEMINAR NASIONAL & CALL FOR PAPER*, 74–79.
- Nurmainia, Fikri, H., Syofiani, Naini, I., & Morelent, Y. (2025). Penggunaan Variasi Bahasa “Slang” pada Kolom Komentar di Platform Media Sosial

- Youtube @Deddy Corbuzier & @Denny Sumargo : Kajian Sociolinguistik. *Jurnal Penelitian Dan Pengkajian Ilmiah*, 2(8), 1425–1426.
- Saputra, A. N. A., Saputro, R. E., & Saputra, D. I. S. (2025). Enhancing Sentiment Analysis Accuracy Using SVM and Slang Word Normalization on YouTube Comments. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 9(2), 687–699.
- Saputro, S. W., Abadi, F., & Nugroho, R. A. (2025). ANALISIS SENTIMEN ULASAN MEDIA SOSIAL UMKM KULINER DENGAN PENDEKATAN LEXICON-BASED DAN KOSAKATA KHUSUS. *JIP (Jurnal Informatika Polinema)*, 12(2), 219–224.
- Sebayang, R. R., Purba, E., Damanik, S. P., & Surip, M. (2024). DINAMIKA BAHASA GAUL DAN SERAPAN ASING DI ERA DIGITAL: DAMPAKNYA TERHADAP KEMAMPUAN BERBAHASA INDONESIA BAKU. *Jurnal Pendidikan Bahasa Dan Sastra*, 05(02), 10–22.
- Sianipar, J. C., Siantar, N. L. L., Barus, E., & Saragih, E. L. (2026). Analisis Variasi dan Jenis Bahasa dalam Perspektif Sociolinguistik pada Konten YouTube Jerome Polin “Jerome Goes to School di SMK Bagimu Negeriku.” *Jurnal Multidisiplin Ilmu Akademik (JMIA)*, 3(2), 27–33.
- Sudirman, Kondolayuk, M. L., Sriwahyuningrum, A., Cahaya, I. M. E., Astuti, N. L. S., Setiawan, J., Tandirerung, W. Y., Rahmi, S., Nusantara, D. O., Indrawati, F., Fitriya, N. L., Aziza, N., Kurniawati, N., Wardhana, A., & Hasanah, T. (2020). *METODOLOGI PENELITIAN 1* (S. Haryanti (ed.)). CV. MEDIA SAINS INDONESIA.
- Sulistyawati, R., Safitri, M. N. D., Sabilla, I. N., & Virnando, A. (2025). Analisis Penggunaan Bahasa Gaul Sebagai Variasi Sosial Oleh Mahasiswa PBSI FKIP UMK Dalam Pembelajaran Sociolinguistik. *Jurnal Bahasa, Sastra, Dan Pendidikan*, 7(2), 429–436.
- Syalsabella, A. P., & Prapanca, A. (2024). Perancangan Website Online Course (KitaPTN)

untuk Persiapan UTBK
Menggunakan Algoritma
Fisher-Yates Shuffle. *JINACS*
(*Journal of Informatics and*
Computer Science), 06, 581–
588.

Ulya, R. H. (2024). Transformasi
Makrolinguistik Bahasa
Indonesia dalam Gamitan
Media Digital : Analisis Wacana
Kritis pada Platform Media
Sosial. *Jurnal Ilmiah Langue*
and Parole, 1(1), 91–99.

Urva, G., Desyanti, Albanna, I.,
Sungkar, M. S., Gunawan, I. M.
A. O., Adhicandra, I.,
Ramadhan, S., Rahardian, R.
L., Herlawati, Handayanto, R.
T., Ariana, A. A. G. B., Hartatik,
Atika, P. D., & Junaidi, S.
(2023). *PENERAPAN DATA*
MINING DI BERBAGAI
BIDANG : Konsep, Metode,
dan Studi Kasus (Efitra &
Sepriano (eds.)). PT. Sonpedia
Publising Indonesia.

Waridah, E. (2008). *EYD & Seputar*
Kebahasa-Indonesiaan (1st
ed.). KawanPustaka