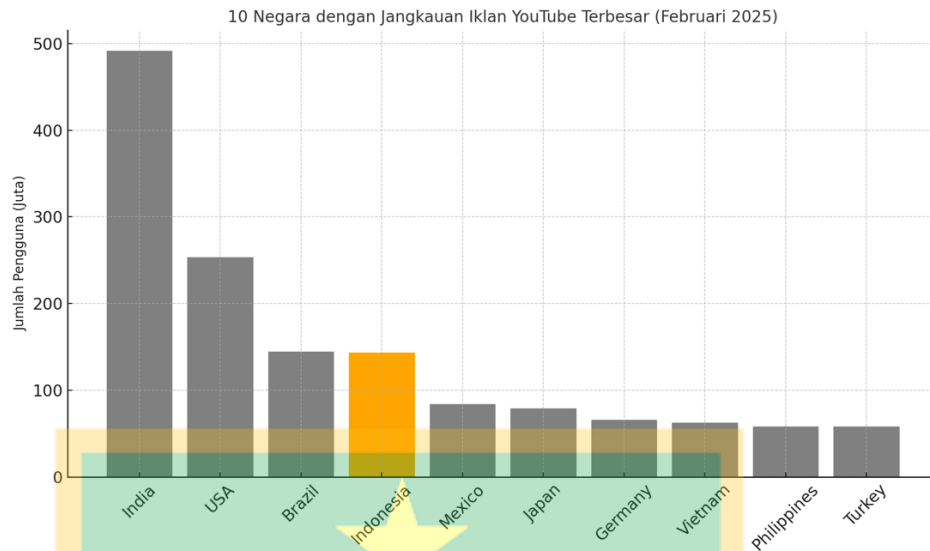


BAB 1 PENDAHULUAN

1.1 Latar Belakang

Sarkasme adalah bentuk komunikasi tidak langsung yang menyampaikan kritik, ejekan, ataupun ketidakpuasan melalui ungkapan yang tampak positif atau netral, namun menyimpan makna berlawanan secara pragmatik. Dalam konteks linguistik, sarkasme sering dikaitkan dengan ironi, hiperbola atau kontras antara maksud dan struktur literal ujaran (Ghosh & Veale, 2017), ini berbeda dari definisi dalam Kamus Besar Bahasa Indonesia yang menekankan aspek ejekan secara eksplisit. Di ranah digital, sarkasme menjadi fenomena umum, terutama dalam komentar di media sosial. Namun, sifat implisit dan multitafsir dari sarkasme menjadikannya tantangan signifikan bagi sistem otomatis, terutama dalam tugas-tugas analisis sentimen, karena dapat membalikkan polaritas makna secara tidak terduga. Oleh karena itu, deteksi sarkasme secara akurat menjadi penting, baik untuk meningkatkan akurasi sistem NLP maupun untuk keperluan moderasi konten dan analisis opini publik.

Pemilihan YouTube sebagai objek penelitian didasarkan pada basis pengguna yang sangat besar di Indonesia. Menurut laporan (We Are Social & Meltwater, 2025), estimasi jangkauan iklan YouTube (*YouTube Ad Reach*) di Indonesia mencapai 143 juta pengguna, atau sekitar 50% dari total populasi nasional. Angka ini menempatkan Indonesia sebagai negara dengan jumlah pengguna YouTube terbesar keempat di dunia sebagaimana ditunjukkan pada Gambar 1.1.



Gambar 1.1. Negara dengan Jangkauan Iklan YouTube Terbesar (Februari 2025)

Sumber: diolah dari (We Are Social & Meltwater, 2025)

Meskipun metrik ini tidak identik dengan jumlah pengguna aktif bulanan, *Ad Reach* digunakan secara luas dalam konteks pemasaran digital sebagai proksi keterjangkauan pengguna aktif.

Hal ini diperkuat oleh data perbandingan jumlah pengguna antar platform media sosial. Masih menurut laporan yang sama, YouTube memiliki jumlah pengguna terbanyak di Indonesia dibandingkan dengan platform sosial media utama lainnya, yaitu 143 juta, mengungguli Facebook (122,3 juta), TikTok (107,7 juta), Instagram (103,4 juta), dan X/Twitter (25,2 juta) sebagaimana ditunjukkan pada Tabel 1.1.

Tabel 1.1. Jumlah Pengguna Media Sosial di Indonesia (Februari 2025)

Platform Sosial Media	Pengguna (juta)
YouTube	143,0
Facebook	122,3
TikTok	107,7
Instagram	103,4
X (twitter)	25,2

Sumber : diolah dari (We Are Social & Meltwater, 2025)

Dengan basis pengguna yang luas dan karakter interaksi yang terbuka, YouTube tidak hanya berperan sebagai media hiburan, tetapi juga menjadi ruang publik digital utama tempat masyarakat menyampaikan opini, kritik, hingga sindiran. Komentar-komentar di platform ini kerap mengandung ekspresi yang eksplisit maupun implisit, termasuk sarkasme, sehingga menjadi sumber data yang sangat representatif untuk penelitian linguistik berbasis teks dalam Bahasa Indonesia.

Dalam bidang pemrosesan bahasa alami (*Natural Language Processing*/NLP), sarkasme merupakan tantangan tersendiri karena maknanya sering kali bertentangan dengan struktur literal kalimat. Deteksi sarkasme sering dikaitkan dengan analisis sentimen karena ekspresi sarkastis dapat menimbulkan ambiguitas makna; pernyataan yang secara sintaktis terlihat positif bisa saja bermakna negatif secara pragmatik.

Jia et al., (2024) menjelaskan bahwa ambiguitas semantik dalam ujaran sarkastis dapat menyebabkan model sentimen gagal mengenali polaritas yang sebenarnya, terutama jika tidak mempertimbangkan konteks pragmatis dan *multimodal*. Hal ini diperkuat oleh temuan Sharma et al., (2022) yang menunjukkan bahwa kehadiran sarkasme secara signifikan menurunkan akurasi sistem analisis sentimen otomatis. Selain itu, Razali et al., (2021) juga menekankan bahwa tanpa integrasi fitur kontekstual, sistem deteksi sarkasme berisiko tinggi melakukan klasifikasi yang keliru terhadap polaritas emosional suatu pernyataan. Namun demikian, dalam konteks tertentu, sarkasme yang eksplisit secara linguistik tetap dapat dikenali melalui pendekatan berbasis teks. Penelitian ini secara sadar membatasi ruang lingkupnya pada jenis sarkasme tersebut, yakni yang mengandung ironi, hiperbola, atau kontras makna yang dapat diamati secara langsung dari bentuk ujaran, tanpa ketergantungan pada data *multimodal*.

Sejumlah studi awal mencoba menghadapi tantangan ini menggunakan pendekatan berbasis fitur manual dan algoritma klasifikasi klasik seperti SVM atau *Random Forest*. Namun, pendekatan konvensional semacam ini memiliki

keterbatasan dalam menangkap makna implisit, terutama jika tanpa pemahaman konteks secara menyeluruh (Ranti & Girsang, 2020; Suhartono et al., 2024).

Seiring perkembangan arsitektur *transformer*, model seperti BERT telah menunjukkan performa yang lebih baik dalam memahami struktur dan konteks kalimat. Islam et al., (2025) menunjukkan bahwa pendekatan *hybrid* BERT-CNN dapat meningkatkan performa klasifikasi komentar YouTube yang bersifat emosional dan opini. Meskipun penelitian tersebut menggunakan data berbahasa Inggris dan tidak secara khusus fokus pada Bahasa Indonesia, pendekatannya tetap relevan sebagai acuan teknis dalam pengembangan sistem deteksi sarkasme.

Upaya awal dalam deteksi sarkasme berbahasa Indonesia dilakukan oleh Suhartono et al., (2024) melalui pengembangan *dataset* IdSarcasm yang bersumber dari platform Twitter dan Reddit. Meskipun signifikan, *dataset* tersebut belum mencerminkan konteks linguistik yang khas dari komentar YouTube. Komentar di YouTube memiliki karakteristik yang khas dibandingkan media sosial lain, misalnya dalam hal panjang teks, gaya bahasa campuran, dan keberagamannya akibat topik video yang ditanggapi. Hal ini menjadikan pendekatan deteksi sarkasme yang efektif pada platform lain belum tentu relevan untuk diterapkan langsung pada komentar YouTube. (Gole et al., 2024), misalnya secara eksplisit menyatakan bahwa temuan mereka yang berbasis dataset Reddit (SARC 2.0) tidak dapat digeneralisasi ke platform lain.

Meskipun penelitian ini tidak mempertimbangkan isi video secara langsung, karakteristik ini tetap menjadikan komentar YouTube sebagai sumber data yang menarik untuk studi deteksi sarkasme berbasis teks. Beberapa penelitian lokal telah memanfaatkan komentar YouTube untuk tugas analisis sentimen. Mandhasiya et al., (2024) mengevaluasi berbagai model *deep learning hybrid* (CNN, LSTM, CNN-LSTM, dan LSTM-CNN) dengan representasi data menggunakan BERT dalam konteks politik Indonesia, dan menemukan bahwa BERT-CNN memberikan akurasi tertinggi dalam mengklasifikasikan komentar positif dan negatif. Sementara itu, Ma'aly et al., (2024) menggunakan pendekatan *multi-label classification* untuk mengidentifikasi emosi seperti kemarahan, antisipasi, jijik, kegembiraan, ketakutan, kesedihan, kejutan dan kepercayaan dari komentar YouTube, dengan

model Bi-LSTM yang mencapai skor AUC sebesar 0,91. Meskipun demikian, kedua studi ini tidak secara spesifik menangani sarkasme sebagai fenomena linguistik yang berbeda. Tabel 1.2 berikut merangkum beberapa studi terdahulu yang menjadi acuan dalam penelitian ini, sekaligus menunjukkan celah penelitian yang ingin diisi.

Tabel 1.2. Perbandingan metode Penelitian Sebelumnya

Penelitian	Bahas	Data	Metode	Fokus	Teknik	Gap
	a	Platform			Labeling	Research
Jia et al. (2024)	Inggris	Twitter	Multimodal, transformer	Sarkasme	Prelabeled	Tidak fokus pada teks
Sharma et al. (2022)	Inggris	Reddit, Twitter, Headline	Hybrid (BERT, LSTM)	Sarkasme	otomatis	Tidak kontekstual pragmatik
Suhartono et al.(2024)	Indonesia	Twitter, Reddit	Fine-tuning IndoBERT	Sarkasme	Prelabeled, otomatis by tag	Belum menyasar Youtube
Mandhasiya et al.(2024)	Indonesia	Youtube	Hybrid BERT VS Deep Learning:CNN-LSTM	Sentimen	Tidak dijelaskan	Tidak membahas sarkasme
Ma'aly et al. (2024)	Indonesia	YouTube	Deep learning : Bi-LSTM, CNN	Emosi	Otomatis dengan GPT-3.5	Tidak eksplisit menangani sarkasme
Gole et all (2024)	Inggris	Reddit SARC 2.0	Perbandingan kinerja model-model GPT (<i>fine-tuning</i> dan <i>zero-shot</i>)	Sarkasme	Prelabeled	Belum menyasar YouTube
Qin et all (2025)	Inggris	Berita online, Reddit	BERT+GGNN	Sarkasme	Prelabeled	Belum menyasar YouTube
Penelitian ini	Indonesia	YouTube	Fine-tuning IndoBERT	Sarkasme eksplisit	Otomatis dengan GPT-4o	Fokus pada pendekatan berbasis teks tanpa multimodal

Berdasarkan celah yang ditinggalkan oleh studi-studi sebelumnya, penelitian ini bertujuan untuk mengeksplorasi deteksi sarkasme dalam komentar YouTube berbahasa Indonesia melalui pendekatan *fine-tuning* terhadap model pralatih IndoBERT. Pemilihan IndoBERT didasarkan pada dua alasan utama. Pertama, IndoBERT IndoNLU merupakan model berbasis *transformer* yang dilatih secara khusus pada korpus besar Bahasa Indonesia (Indo4B) yang mencakup ragam formal dan informal dari berbagai domain, termasuk media sosial, berita, dan blog. Kedua, menurut benchmark IndoNLU Wilie et al., (2020), IndoBERT secara konsisten melampaui performa model multibahasa seperti mBERT dan XLM-R dalam berbagai tugas klasifikasi dan ekstraksi informasi berbahasa Indonesia. Oleh karena itu, IndoBERT dinilai lebih sesuai secara linguistik dan empiris untuk menangani tugas klasifikasi teks dengan nuansa lokal seperti deteksi sarkasme dalam komentar YouTube.

Penelitian ini memilih untuk membangun *dataset* baru yang secara khusus merepresentasikan karakteristik komentar YouTube berbahasa Indonesia, yang memiliki gaya komunikasi, ekspresi ironi, dan konteks budaya yang berbeda dari data sosial media berbahasa Inggris seperti Reddit atau Twitter.

Perkembangan model bahasa besar (*large language models/LLM*) seperti GPT membuka peluang baru dalam pemrosesan bahasa alami, termasuk dalam tugas pelabelan data. Dataset utama dalam penelitian ini disusun melalui proses *scraping* komentar YouTube berbahasa Indonesia dari video dengan topik yang berpotensi memunculkan respons sarkastis, seperti politik, isu publik, dan komedi. Pelabelan data dilakukan sepenuhnya secara otomatis menggunakan model bahasa GPT-4o, berdasarkan *prompt* sistem yang dirancang khusus untuk mengidentifikasi komentar sarkasme. *Prompt* ini memuat definisi operasional sarkasme dan indikator linguistik eksplisit, seperti ironi, hiperbola, dan pujian terselubung, sehingga model dapat mengklasifikasikan komentar menjadi “sarkasme” atau “non-sarkasme” secara konsisten.

Penggunaan model bahasa besar (LLM) dalam pelabelan otomatis telah dicoba pada berbagai studi sebelumnya. Ma’aly et al., (2024), misalnya, memanfaatkan GPT-3.5 untuk anotasi emosi pada komentar YouTube secara

otomatis. Gole et al., (2024) menemukan bahwa deteksi otomatis sarkasme menggunakan GPT-3.5 Turbo memiliki keterbatasan dalam mengenali sarkasme dibandingkan model yang telah di-*fine-tune*. Qin et al., (2025) juga menggarisbawahi bahwa LLM sering menghadapi kesulitan mendeteksi sarkasme halus karena kecenderungan mengandalkan pola probabilistik, sehingga berisiko melewati nuansa ironi dan kontras pragmatis.

Berdasarkan temuan-temuan tersebut, penelitian ini merancang strategi pelabelan otomatis yang mengandalkan GPT-4o, dengan penekanan pada perancangan prompt sebagai kunci kualitas label. Pendekatan ini dipilih untuk memaksimalkan efisiensi waktu dan sumber daya, sekaligus menjaga konsistensi pelabelan tanpa intervensi manual. Dengan metode ini, seluruh *dataset* terbentuk dari hasil klasifikasi otomatis GPT-4o sesuai pedoman linguistik yang telah ditetapkan.

Lebih lanjut, strategi *fine-tuning* yang digunakan dalam penelitian ini juga diperkuat oleh temuan (Gole et al., 2024) yang menunjukkan bahwa model GPT-3 *davinci* yang di-*fine-tune* secara khusus pada data sarkasme berhasil mengungguli pendekatan *zero-shot* GPT-4 dalam hal akurasi dan F1-score (0.81 vs 0.75). Meskipun studi tersebut menggunakan *dataset* SARC 2.0 yang berbahasa Inggris, hasil tersebut menunjukkan bahwa *fine-tuning* dapat secara signifikan meningkatkan sensitivitas model terhadap ekspresi sarkastis, dibandingkan pendekatan tanpa pelatihan tambahan. Prinsip ini menjadi salah satu justifikasi teknis pemilihan metode *fine-tuning* terhadap IndoBERT dalam penelitian ini.

Pendekatan pelabelan otomatis ini diharapkan mampu memperluas cakupan data secara efisien sambil tetap mempertahankan konsistensi klasifikasi. Penelitian ini juga memberi perhatian khusus terhadap tantangan metodologis dalam pelabelan sarkasme, yang secara konseptual berbeda dari sentimen biasa karena melibatkan ambiguitas, konteks pragmatik, dan potensi multitafsir makna.

Dengan demikian, penelitian ini tidak hanya bertujuan mengembangkan model klasifikasi yang lebih peka terhadap bentuk-bentuk ironi dalam Bahasa Indonesia, tetapi juga berkontribusi terhadap pengayaan literatur NLP lokal, khususnya dalam konteks deteksi sarkasme pada platform YouTube. Hasil

penelitian ini diharapkan menjadi fondasi bagi pengembangan sistem moderasi otomatis dan analisis opini publik yang lebih adaptif terhadap gaya komunikasi digital masyarakat Indonesia.

1.2 Rumusan Masalah

1. Bagaimana mengembangkan model deteksi sarkasme yang efektif pada komentar YouTube berbahasa Indonesia menggunakan pendekatan *fine-tuning* IndoBERT?
2. Bagaimana efektivitas strategi mitigasi ketidakseimbangan kelas dalam meningkatkan performa model deteksi sarkasme?
3. Varian IndoBERT manakah yang paling optimal dan efektif untuk tugas deteksi sarkasme pada komentar YouTube berbahasa Indonesia

1.3 Tujuan

1. Mengembangkan model deteksi sarkasme yang efektif untuk komentar YouTube berbahasa Indonesia menggunakan pendekatan *fine-tuning* terhadap model prelatih IndoBERT
2. Mengevaluasi efektivitas berbagai strategi mitigasi ketidakseimbangan kelas (*imbalanced*, *under sampling* dan *class weighting*) dalam meningkatkan performa model.
3. Mengevaluasi dan membandingkan kinerja model varian IndoBERT dalam mengklasifikasikan komentar sarkastis dan non-sarkastis menggunakan metrik evaluasi kuantitatif, seperti akurasi, *precision*, *recall* dan *F1-score*.

1.4 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Memberikan kontribusi akademik dalam pengembangan sistem deteksi sarkasme berbasis pemrosesan bahasa alami dalam Bahasa Indonesia.
2. Menyediakan *dataset* sarkasme komentar YouTube berbahasa Indonesia yang telah dilabeli dengan GPT-4o.
3. Menjadi referensi awal bagi pengembangan sistem moderasi komentar otomatis berbasis kecerdasan buatan pada platform YouTube.

4. Memberikan insight tentang efektivitas *pre-training objectives* yang berbeda (IndoNLU vs IndoLEM) untuk tugas deteksi sarkasme dalam bahasa Indonesia informal

1.5 Batasan Masalah

Untuk menjaga agar ruang lingkup penelitian tetap fokus dan terukur, maka penelitian ini dibatasi pada hal-hal sebagai berikut :

1. Komentar yang dianalisis terbatas pada Bahasa Indonesia dari platform YouTube.
2. Data dikumpulkan melalui proses *scraping* dari video yang berpotensi memunculkan komentar sarkastis, seperti konten bertema politik, isu publik, militer atau komedi.
3. Pelabelan data dilakukan secara otomatis sepenuhnya menggunakan bantuan model GPT-4o untuk anotasi, berdasarkan sistem *prompt* yang dirancang khusus untuk mendeteksi sarkasme.
4. Penelitian ini hanya menganalisis konten teks komentar tanpa mempertimbangkan isi video, meta data, utas percakapan, atau elemen *multimodal* lainnya.
5. Model utama yang digunakan adalah IndoBERT-base (IndoNLU), dengan perbandingan terhadap varian model lain yakni IndoBERT-base IndoLEM.

1.6 Sistematika Pembahasan

Sistematika pembahasan dalam tesis ini disusun sebagai berikut :

Bab 1 Pendahuluan, menjelaskan latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, dan sistematika penulisan.

Bab 2 Landasan kepastakaan , memuat teori-teori yang relevan, termasuk pengertian sarkasme, analisis sentimen, pemrosesan bahasa alami (NLP), model *transformer*, IndoBERT, serta hasil penelitian terdahulu dan kerangka konseptual.

Bab 3 Metodologi Penelitian, menjabarkan tahapan teknis, mulai dari pengumpulan data komentar YouTube, teknik pelabelan, *preprocessing* data, strategi *fine-tuning* IndoBERT, hingga evaluasi model dan eksperimen pembandingan dengan varian model IndoBERT lain (IndoLEM).

Berdasarkan uraian pada bab ini, penelitian diarahkan untuk menjawab rumusan masalah melalui pendekatan yang dibatasi secara metodologis dan konteks data. Untuk mendukung pemilihan pendekatan dan merancang metodologi yang tepat, bab selanjutnya akan membahas landasan teoritis dan hasil-hasil penelitian terdahulu yang relevan.

Sebagai pelengkap, penelitian ini juga menyertakan eksperimen pembandingan dengan varian lain dari IndoBERT, yakni IndoLEM, untuk mengevaluasi pengaruh sumber korpus *pre-training* terhadap performa deteksi sarkasme.

