



Implementasi Model IndoBERT untuk Klasifikasi Komentar Berbahasa Indonesia pada Video Review Smartphone di YouTube

Ria Natali Alfarizi *, Saifu Rohman, Nur Hasanah

Program studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Sains Al-Qur'an, Indonesia

*Email (Penulis Korespondensi): rianalfarizi321@gmail.com

Abstrak. YouTube menjadi salah satu sumber referensi utama masyarakat Indonesia dalam mencari informasi mengenai smartphone sebelum melakukan pembelian, salah satunya melalui kolom komentar pada video review. Namun, tidak semua komentar mengandung informasi yang relevan, sehingga menyulitkan pengguna dalam menemukan komentar yang benar-benar informatif. Penelitian ini membangun sistem klasifikasi komentar otomatis menggunakan model IndoBERT untuk mengelompokkan komentar ke dalam kategori Informatif dan Non-Informatif, serta pengelompokan sub-kategori topik menggunakan pendekatan berbasis kata kunci. Data komentar dikumpulkan dari video review smartphone berbahasa Indonesia di YouTube, melalui tahapan pra-pemrosesan meliputi pembersihan teks, normalisasi, tokenizing, dan pelabelan data, sebelum model IndoBERT di-fine-tune untuk memahami konteks bahasa Indonesia. Hasil pengujian menunjukkan model IndoBERT mencapai akurasi 84,33%, presisi 83,66%, recall 85,33%, dan F1-score 84,49%, yang menunjukkan bahwa model ini efektif dalam memahami konteks komentar berbahasa Indonesia serta membedakan komentar informatif dan non-informatif secara akurat. Sistem ini diimplementasikan sebagai aplikasi web yang diharapkan dapat membantu pengguna menyaring informasi relevan dari kolom komentar YouTube secara lebih efisien.

Kata kunci: IndoBERT; Klasifikasi Teks; Komentar YouTube; Komentar Informatif; Review Smartphone

Abstract. YouTube has become a primary source of information for Indonesians researching smartphones before making a purchase, particularly through the comment sections of review videos. However, not all comments contain relevant information, making it difficult for users to identify those that are truly informative. This study developed an automated comment classification system using the IndoBERT model to categorize comments as either "Informative" or "Non-Informative," while also grouping them into topic sub-categories using a keyword-based approach. Comment data was collected from Indonesian-language smartphone review videos on YouTube and underwent preprocessing – including text cleaning, normalization, tokenization, and labeling – before the IndoBERT model was fine-tuned to comprehend the Indonesian language context. Test results demonstrated that the IndoBERT model achieved an accuracy of 84.33%, precision of 83.66%, recall of 85.33%, and an F1-score of 84.49%, indicating its effectiveness in understanding the context of Indonesian comments and accurately distinguishing between informative and non-informative ones. The system was implemented as a web application designed to help users more efficiently filter relevant information from YouTube comment sections.

Keywords: IndoBERT; Text Classification; YouTube Comments; Informative Comments; Smartphone Reviews

1. Pendahuluan

Perkembangan teknologi *smartphone* yang sangat pesat telah mendorong meningkatnya konsumsi konten ulasan produk pada platform video, khususnya YouTube. Di Indonesia, video review *smartphone* semakin sering dijadikan rujukan oleh calon konsumen sebelum melakukan pembelian, sementara kolom komentar menghadirkan pengalaman langsung dari pengguna yang dapat memperkaya pemahaman terhadap kinerja produk dalam penggunaan nyata (Mefdi & Halim, 2025); (Cahyani et al., 2025). Dalam konteks perilaku konsumen digital, komentar pengguna berpotensi memberikan informasi yang lebih beragam dibandingkan narasi *reviewer*, karena memuat pengalaman pemakaian jangka panjang, perbandingan antar model, serta respons spontan terhadap kelebihan dan kekurangan produk (Mulyono et al., 2024); (Kamilah et al., 2025).

Meskipun demikian, tidak seluruh komentar pada video review mengandung informasi yang relevan. Sebagian komentar justru berupa ekspresi singkat, spam, emoji, promosi, atau tanggapan yang tidak berkaitan langsung dengan produk, sehingga proses identifikasi komentar bernilai menjadi tidak efisien apabila dilakukan secara manual (Monica & Purwanto, 2026). Kondisi ini diperburuk oleh keterbatasan fitur bawaan platform yang belum menyediakan penyaringan komentar berdasarkan tingkat informatif, sehingga pengguna tetap harus menelusuri banyak komentar untuk menemukan opini yang benar-benar bermanfaat (Cahyani et al., 2025). Oleh karena itu, dibutuhkan sistem otomatis yang mampu memilah komentar berdasarkan nilai informasinya agar proses ekstraksi wawasan menjadi lebih cepat dan terarah.

Karakter bahasa pada komentar YouTube juga menimbulkan tantangan tersendiri dalam pengolahan teks. Komentar berbahasa Indonesia di media sosial umumnya menggunakan ragam informal, singkatan, istilah gaul, variasi lokal, dan *code-mixing* antara bahasa Indonesia dan bahasa Inggris, yang menyulitkan proses pemahaman teks secara konvensional (Karim & Wibowo, 2025); (Nurjoko & Rahardi, 2024); (Sumartha, 2025). Sejumlah penelitian menunjukkan bahwa *Support Vector Machine* (SVM) tetap mampu memberikan hasil yang baik dalam berbagai tugas klasifikasi teks berbahasa Indonesia (Salsabilla & Nunik Pratiwi, 2024). Namun, pendekatan ini masih sangat bergantung pada representasi fitur statistik seperti TF-IDF, sehingga kemampuan menangkap konteks semantik yang lebih kompleks menjadi terbatas pada data yang sangat bervariasi dan tidak formal (Salsabilla & Nunik Pratiwi, 2024). Oleh karena itu, model berbasis konteks seperti IndoBERT dinilai lebih sesuai untuk menangani karakter komentar YouTube yang dinamis, karena mampu memahami relasi antarkata secara bidireksional dan lebih adaptif terhadap ekspresi informal pengguna (Mahdi Faza & Moch. Taufik, 2025).

Sebagai alternatif, model berbasis *Transformer* seperti IndoBERT menawarkan kemampuan representasi kontekstual yang lebih baik karena memahami kata secara bidireksional dalam suatu kalimat. Sejumlah penelitian menunjukkan bahwa *fine-tuning* IndoBERT memberikan kinerja yang lebih unggul pada berbagai tugas pemrosesan bahasa Indonesia, termasuk klasifikasi teks dan analisis sentimen (Nuryadi et al., 2025); (Pranata et al., 2024). Keunggulan ini menjadi dasar pemilihan IndoBERT untuk menangani komentar YouTube yang mengandung bahasa informal dan *code-mixing*, karena model tersebut lebih adaptif dalam menangkap makna tersirat, konteks lokal, dan variasi ekspresi pengguna dibandingkan metode berbasis fitur statistik (Nurjoko & Rahardi, 2024); (Simarmata & Sasongko, 2025).

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan sistem klasifikasi komentar video review *smartphone* berbahasa Indonesia menggunakan *fine-tuning* IndoBERT dengan dua kelas utama, yaitu informatif dan non-informatif. Komentar yang telah terklasifikasi akan dipetakan lebih lanjut ke dalam 16 sub-kategori berbasis kata kunci, dan sistem akan diimplementasikan dalam bentuk aplikasi web agar dapat digunakan secara praktis oleh pengguna (Alvin et al., 2025); (Simanjuntak et al., 2024). Kebaruan penelitian ini terletak pada penggabungan *fine-tuning* IndoBERT dua kelas dengan penerapan ambang kepercayaan (*confidence threshold*) 65% pada kelas Informatif untuk menjaga presisi label akhir, pemetaan sub-kategori yang lebih rinci, dan implementasi web untuk membantu calon konsumen memperoleh informasi yang relevan secara lebih efisien. Secara tidak langsung, penelitian ini turut mendukung capaian *Sustainable Development Goals* (SDGs), khususnya SDG 9 (Industri, Inovasi, dan Infrastruktur) melalui pemanfaatan model bahasa berbasis Transformer untuk menyaring informasi konsumen secara otomatis, serta SDG 12 (Konsumsi dan Produksi yang Bertanggung Jawab) dengan membantu calon konsumen memperoleh informasi produk yang lebih relevan dan berbasis data sebelum melakukan pembelian.

2. Metode

Penelitian ini menerapkan pendekatan kuantitatif, di mana data komentar YouTube diolah dan direpresentasikan dalam bentuk numerik untuk dianalisis menggunakan model pembelajaran mesin, sehingga kinerja sistem dapat diukur secara objektif melalui metrik evaluasi yang terukur dan dapat diuji ulang. Pengembangan sistem dilakukan dengan metode prototyping, dimulai dari pengumpulan data komentar video review *smartphone* di YouTube, dilanjutkan dengan pelabelan dan pra-pemrosesan data. Data yang telah siap kemudian dibagi dan digunakan untuk melakukan *fine-tuning* model IndoBERT, sebelum akhirnya kinerja model dievaluasi menggunakan *confusion matrix*.

2.1. Pengumpulan Data

Tahap awal penelitian ini adalah mengumpulkan data komentar dari sejumlah video *review smartphone* berbahasa Indonesia di YouTube menggunakan YouTube Data API v3. Setiap komentar yang terkumpul mencakup teks komentar beserta atribut pendukung seperti nama pengguna dan jumlah likes. Proses ini menghasilkan kumpulan data mentah yang selanjutnya digunakan sebagai dasar pembentukan dataset penelitian.

2.2. Pelabelan Data

Setelah data komentar terkumpul, tahap selanjutnya adalah pemberian label secara manual terhadap setiap komentar ke dalam dua kelas, yaitu Informatif (dilabeli 1) dan Non-Informatif (dilabeli 0), berdasarkan ada tidaknya informasi yang bermanfaat terkait produk di dalam komentar tersebut. Proses pelabelan dilakukan dengan memperhatikan keseimbangan jumlah data pada kedua kelas, sehingga diperoleh dataset akhir sebanyak 1.500 komentar yang terdiri atas 750 komentar Informatif dan 750 komentar Non-Informatif. Proses pelabelan dilakukan secara manual oleh satu orang pelabel, yaitu peneliti sendiri, berdasarkan definisi operasional yang telah ditetapkan sebelumnya: komentar dikategorikan Informatif apabila memenuhi minimal satu dari empat kriteria, yaitu memuat pengalaman penggunaan nyata, menyertakan data atau spesifikasi teknis, bersifat objektif dengan menyebutkan kelebihan dan kekurangan produk secara seimbang, atau relevan dengan

kebutuhan calon pembeli; sedangkan komentar yang hanya berisi ekspresi singkat, spam, atau promosi tanpa memenuhi kriteria tersebut dikategorikan Non-Informatif. Karena pelabelan dilakukan oleh satu pelabel, penelitian ini tidak melaporkan ukuran kesepakatan antarpenilai (inter-annotator agreement) seperti Cohen's Kappa, sehingga potensi subjektivitas dalam penentuan label menjadi salah satu keterbatasan penelitian ini yang perlu diperhatikan pada penelitian selanjutnya, misalnya dengan melibatkan lebih dari satu pelabel independen.

2.3. Pra-pemrosesan Data

Data yang telah diberi label selanjutnya melalui tahap pra-pemrosesan untuk mempersiapkan teks sebelum diproses oleh model. Tahap ini meliputi:

- Cleaning*, yaitu menghapus tanda baca, karakter spesial, angka yang tidak relevan, serta spasi berlebih pada teks komentar.
- Normalisasi*, yaitu mengubah kata tidak baku atau singkatan yang umum digunakan dalam komentar media sosial ke bentuk kata baku agar lebih mudah dipahami model.
- Case folding*, yaitu menyeragamkan seluruh huruf pada teks menjadi huruf kecil untuk mengurangi variasi yang tidak diperlukan.
- Tokenizing*, yaitu memecah teks komentar menjadi unit-unit token menggunakan *tokenizer* bawaan IndoBERT, sehingga teks siap diproses pada tahap pelatihan model.

2.4. Pembagian Data

Data yang telah melalui pra-pemrosesan selanjutnya dibagi menjadi data latih dan data uji dengan rasio 80:20, sehingga diperoleh 1.200 data latih dan 300 data uji. Data uji ini dipisahkan sejak awal dan hanya digunakan pada tahap evaluasi akhir model. Sebagian dari 1.200 data latih tersebut kemudian disisihkan sebagai data validasi dengan rasio 90:10 (1.080 data latih aktif dan 120 data validasi), yang performanya turut dipantau pada tiap epoch untuk mendeteksi indikasi overfitting selama proses pelatihan berlangsung.

2.5. Proses Training dan Fine-Tuning Model IndoBERT

Pada tahap ini, model IndoBERT dilatih menggunakan data latih yang telah diproses untuk mempelajari konteks dan pola bahasa pada komentar berbahasa Indonesia. Mekanisme utama dalam arsitektur ini adalah *scaled dot-product attention*, yang dirumuskan pada persamaan (1).

$$Q, K, V = QKTdk \ V \quad (1)$$

dengan Q (*query*), K (*key*), dan V (*value*) merupakan representasi *vektor* dari *input* teks, sedangkan d_k adalah dimensi *vektor key* yang berperan sebagai faktor penyekalaan agar nilai *dot-product* tidak terlalu besar. Representasi vektor dari token [CLS] hasil keluaran IndoBERT selanjutnya diteruskan ke lapisan klasifikasi (fully connected layer) untuk menghasilkan probabilitas kelas menggunakan fungsi *softmax* seperti pada persamaan (2).

$$y_i | x = \text{ezije}zj(2)$$

dengan z_i merupakan nilai logit untuk kelas ke- i , dan $P(y_i | x)$ adalah probabilitas komentar x termasuk ke dalam kelas y_i . Proses pelatihan dilakukan dengan memperbarui bobot model pra-latih menggunakan fungsi *loss cross-entropy* yang dirumuskan pada persamaan (3).

$$L = -iy \log y_i \quad (3)$$

dengan y_i adalah label aktual dan $\hat{y}_i$ adalah probabilitas prediksi model untuk kelas tersebut, yang dioptimasi menggunakan algoritma Adam dengan learning rate sebesar $3e-5$ selama maksimal 10 *epoch*, disertai penerapan *early stopping* apabila tidak terdapat peningkatan performa validasi secara signifikan.

2.6. Pengelompokan Sub-kategori

Komentar yang telah terklasifikasi ke dalam label Informatif dan Non-Informatif selanjutnya dikelompokkan lebih lanjut ke dalam sub-kategori menggunakan pendekatan berbasis kata kunci (*keyword-based*). Pendekatan ini bekerja dengan cara mencocokkan kata-kata yang terdapat dalam teks komentar terhadap daftar kata kunci yang telah ditetapkan untuk masing-masing sub-kategori. Apabila sebuah komentar mengandung kata kunci yang sesuai, maka komentar tersebut akan dikelompokkan ke dalam sub-kategori yang berkaitan. Jika tidak ditemukan kata kunci yang cocok dari sub-kategori manapun, komentar akan diberi label "Lainnya".

Secara keseluruhan terdapat 16 sub-kategori yang dibagi berdasarkan label utamanya. Label Informatif memiliki 9 sub-kategori yang merepresentasikan aspek-aspek teknis dan fungsional smartphone yang lazim dibahas dalam ulasan pengguna, yaitu Desain (kata kunci: bodi, warna, material, tipis, kokoh), Layar (resolusi, refresh rate, AMOLED, smooth, cerah), Kamera (foto, video, zoom, night mode, buram), Performa (chipset, gaming, fps, benchmark, lag), Baterai (fast charging, daya tahan, boros, awet), UI/Software (Android, bug, bloatware, update, lemot), Audio (speaker, stereo, bass, kencang, pecah), Harga (mahal, murah, worth it, budget, harga), dan Isu/Kekurangan (overheat, bootloop, error, hang, panas).

Label Non-Informatif memiliki 7 sub-kategori yang merepresentasikan jenis komentar yang tidak memberikan informasi substantif, yaitu Pujian (mantap, keren, the best, bagus), Request (bang bahas, next video, review juga, request), Humor (wkwk, haha, lol, ngakak), Promosi (shopee, lazada, subscribe, promo, diskon), Reaksi Emosional (pengen, ngiri, mupeng, impian), Opini Tanpa Alasan (jelek, skip, gak suka, ga worth), dan Pertanyaan Tidak Relevan (lagu apa, nomor wa, ig-nya, link).

2.7. Evaluasi Model

Kinerja model diukur menggunakan *confusion matrix* dan sejumlah metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

- Accuracy* mengukur perbandingan antara jumlah prediksi yang benar dengan keseluruhan jumlah prediksi yang dilakukan model.
- Precision* mengukur proporsi prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dihasilkan model.
- Recall* mengukur proporsi prediksi positif yang benar dibandingkan dengan seluruh data yang sebenarnya berlabel positif.

d. *F1-score* merupakan rata-rata harmonis antara precision dan recall yang menggambarkan keseimbangan performa model secara keseluruhan.

3. Hasil dan Pembahasan

3.1. Hasil Pengumpulan Data

Data pada penelitian ini dikumpulkan melalui proses scraping komentar YouTube menggunakan YouTube Data API v3 secara otomatis berdasarkan tautan video review smartphone berbahasa Indonesia yang dimasukkan ke dalam sistem. Proses ini menghasilkan data mentah berupa komentar yang belum melalui tahap seleksi maupun pembersihan, mencakup informasi seperti nama penulis komentar, isi teks komentar, tanggal publikasi, dan jumlah likes pada masing-masing komentar. Seluruh contoh data pada Tabel 1 berasal dari video yang sama (ID video: ZGbp60qeIUM), sehingga kolom video_id tidak ditampilkan berulang pada tabel. Contoh hasil scraping data mentah komentar YouTube disajikan pada Tabel 1.

Tabel 1. Hasil Pengumpulan Data

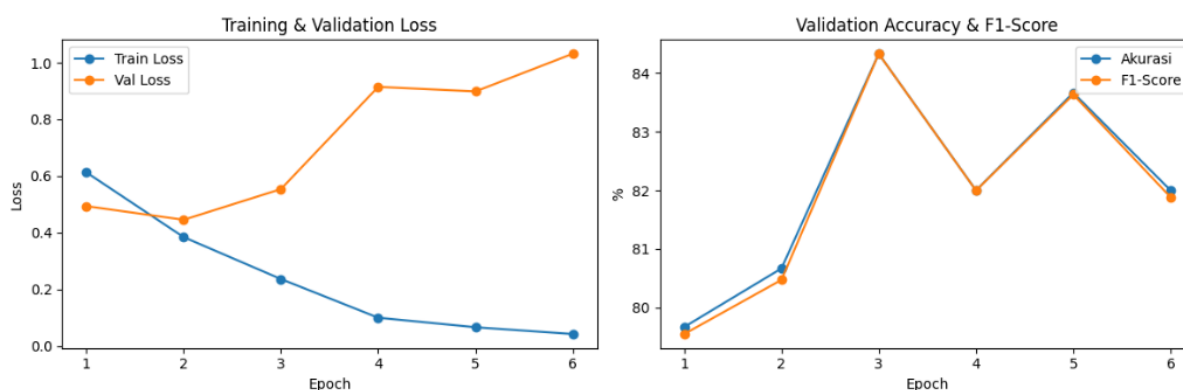
video_id	komentar	penulis	likes	tanggal	label
ZGbp60qeIUM	Tar jg nongol lagi A58, A59 , A60 A100 🤔 kaga ada abisnya. Pake yg ada aja syukuri. Kecuali emang udh bener-bener harus ganti	@Bangs elonn	3345	2026-04-09T03:13:31Z	
ZGbp60qeIUM	Dengerin upgradenya kita ngantuk, dengerin harganya kita meleak wkwk	@IbnuFarhann	3023	2026-04-09T03:03:14Z	
ZGbp60qeIUM	Conclusion: Samsung A57 adalah A56 yang dijual di tahun 2026 yang cuma ganti baju.	@J-Contans	1439	2026-04-09T02:48:21Z	
ZGbp60qeIUM	Aku maaih pake a50s 6GB, 128GB. Beli Oktober 2019 masih oke2 aja nih, udah kecemplung ke selokan kebanting beberapa kali, masih kenceng nih. Mantep emang SAMSUNG	@derite nabe742	7	2026-04-24T10:43:23Z	
ZGbp60qeIUM	"Yup OVERPRICE 🤔 lawannya udah pada punya lensa tele zoom padahal ini gak, the real memanfaatkan kelangkaan ram 💧 wkwkwk, gw pake samsung tapi kalo kalian mau ganti hp tahun ini gw masih rekomendasi harga sekitar ini atau dibawah nya. S25 fe / s24 fe atau a56 / a36 / a55 cari di online semua...	@jiqballl	707	2026-04-09T02:41:52Z	

Berdasarkan Tabel 1, data mentah yang diperoleh masih mengandung berbagai bentuk teks yang tidak terstruktur, seperti penggunaan bahasa informal, singkatan, simbol, dan percampuran bahasa Indonesia dengan bahasa Inggris (*code-mixing*). Oleh karena itu, data tersebut perlu melalui tahap praproses sebelum dapat digunakan untuk melatih model klasifikasi. Dari keseluruhan data yang berhasil dikumpulkan, sebanyak 1.500 komentar dipilih dan diberi label secara manual ke dalam kelas Informatif (750 komentar) dan Non-Informatif (750 komentar) untuk dijadikan dataset pelatihan model.

3.2. Penerapan Model IndoBERT

Penerapan model IndoBERT dalam penelitian ini dilakukan untuk mengklasifikasikan komentar YouTube berbahasa Indonesia pada video review smartphone ke dalam kelas Informatif dan Non-Informatif. Proses *fine-tuning* dilakukan menggunakan Google Colaboratory dengan GPU T4, bahasa pemrograman Python, dan framework PyTorch. Tahap awal yang dilakukan adalah praproses data, mencakup pembersihan teks dari karakter tidak relevan, normalisasi kata tidak baku, case folding, serta tokenisasi menggunakan tokenizer bawaan IndoBERT. Dataset yang digunakan terdiri dari 1500 komentar yakni 750 komentar informatif dan 750 komentar non-informatif, kemudian dibagi menjadi 1.200 data latih dan 300 data uji untuk membangun dan mengevaluasi model klasifikasi.

Proses *fine-tuning* direncanakan selama 10 *epoch* dengan learning rate sebesar $3e-5$ dan *optimizer Adam*. Namun, proses pelatihan berhenti lebih awal pada *epoch* ke-6 karena tidak terdapat peningkatan performa yang signifikan pada data validasi. Grafik *training & validation loss* serta akurasi dan *F1-Score* per *epoch* disajikan pada Gambar 1. Berdasarkan grafik tersebut, *train loss* menunjukkan penurunan yang konsisten seiring bertambahnya *epoch*, sementara *validation loss* cenderung meningkat setelah *epoch* ke-3, yang mengindikasikan terjadinya *overfitting* pada *epoch-epoch* berikutnya. Nilai akurasi dan *F1-Score* validasi tertinggi diperoleh pada *epoch* ke-3, sehingga bobot model pada *epoch* tersebut dipilih sebagai model final.



Gambar 1. Training & Validation Loss dan Validation Accuracy & F1-Score

3.3. Evaluasi Model IndoBERT

Evaluasi model dilakukan menggunakan 300 data uji dari total 1.500 data yang dikumpulkan, yang terdiri atas 150 data kelas Informatif dan 150 data kelas Non-Informatif. Hasil evaluasi menunjukkan bahwa model IndoBERT menghasilkan akurasi sebesar 84,33%, presisi 83,66%, recall 85,33%, dan F1-Score 84,49%. Secara keseluruhan, model mampu

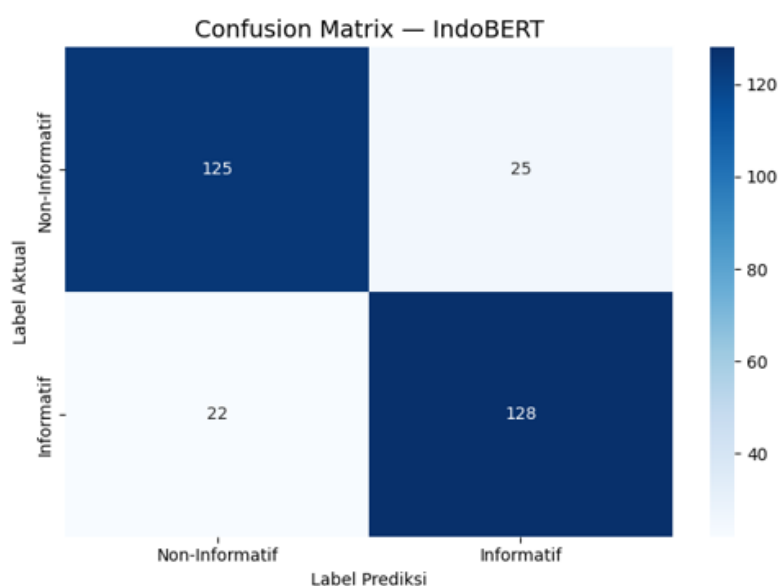
membedakan komentar Informatif dan Non-Informatif dengan baik, dengan performa yang seimbang pada kedua kelas. Hasil lengkap evaluasi model disajikan pada Tabel 2.

Tabel 2. Confusion Matrix

Kelas	Precision	Recall	F1-Score	Support
Non-Informatif	0.85	0.83	0.84	150
Informatif	0.84	0.85	0.84	150
Accuracy			0.84	300
Macro Avg	0.84	0.84	0.84	300
Weighted Avg	0.84	0.84	0.84	300

Berdasarkan Tabel 2, kelas Non-Informatif memperoleh presisi sebesar 0,85 dan recall 0,83, sedangkan kelas Informatif memperoleh presisi 0,84 dan recall 0,85. Performa kedua kelas yang relatif seimbang ini dipengaruhi oleh distribusi data yang seimbang pada masing-masing kelas (150 data per kelas). Nilai macro average dan weighted average yang sama, yaitu 0,84 pada seluruh metrik, mengkonfirmasi bahwa model tidak condong (bias) terhadap salah satu kelas.

Hasil klasifikasi model divisualisasikan menggunakan confusion matrix pada Gambar 2. Berdasarkan confusion matrix tersebut, dari 150 data kelas Non-Informatif, sebanyak 125 data terklasifikasi dengan benar dan 25 data terklasifikasi sebagai Informatif. Dari 150 data kelas Informatif, sebanyak 128 data terklasifikasi dengan benar dan 22 data terklasifikasi sebagai Non-Informatif.



Gambar 2. Confusion Matrix

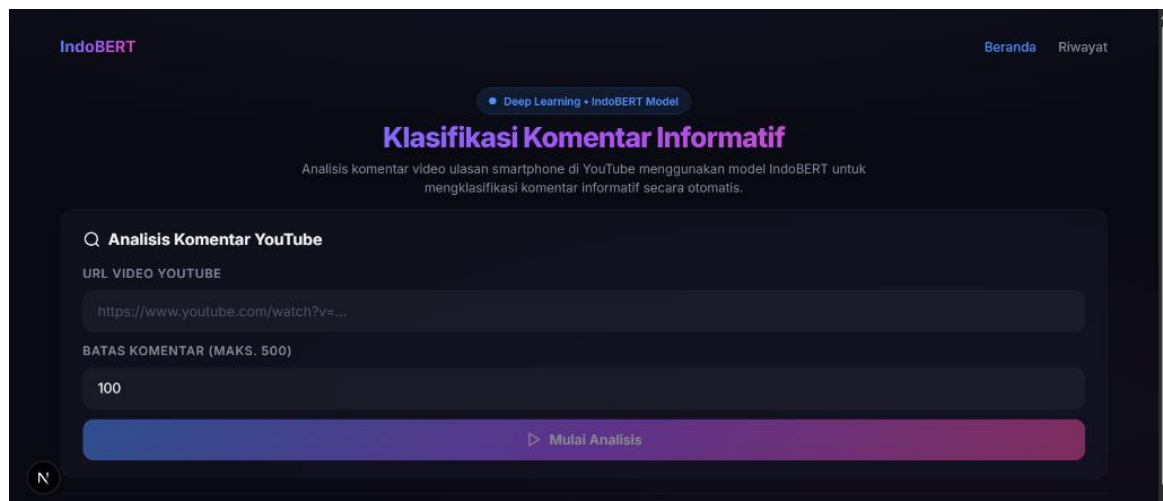
3.4. Implementasi Sistem

Sistem klasifikasi komentar YouTube dirancang dan diimplementasikan dalam bentuk aplikasi web menggunakan framework Flask dengan bahasa pemrograman Python dan basis data Supabase (PostgreSQL). Antarmuka sistem dirancang agar mudah digunakan, sehingga pengguna cukup memasukkan tautan video YouTube dan sistem secara otomatis akan mengekstraksi komentar menggunakan YouTube Data API v3, mengklasifikasikannya

menggunakan model IndoBERT, mengelompokkan sub-kategori menggunakan pendekatan *keyword-based*, serta menampilkan hasil beserta statistik visualisasinya. Tampilan antarmuka sistem akan disajikan sebagai berikut.

a. Tampilan Beranda

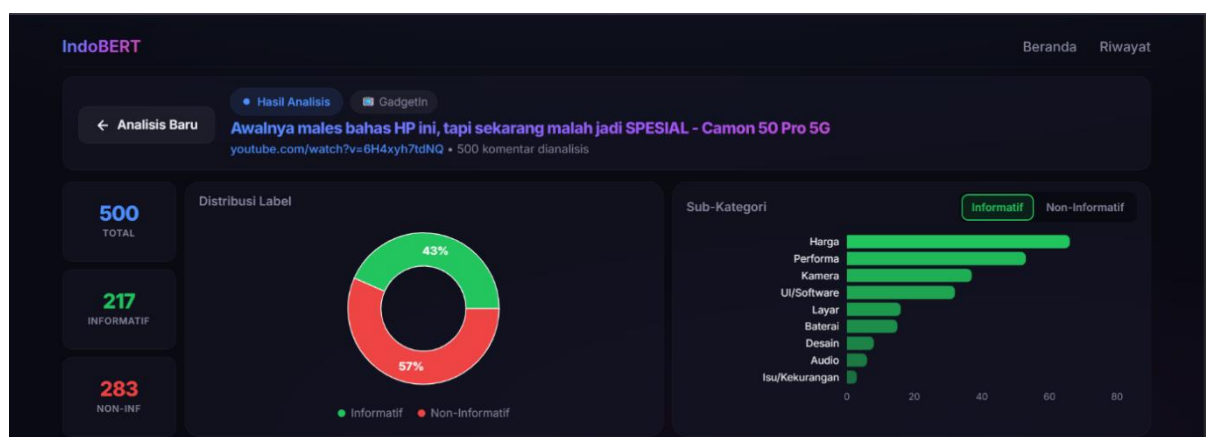
Berfungsi sebagai halaman awal tempat pengguna memasukkan tautan video YouTube dan menentukan jumlah komentar yang ingin dianalisis, sebagai titik masuk sebelum sistem menjalankan proses ekstraksi dan klasifikasi. Pada halaman ini, pengguna cukup menempelkan tautan video ulasan smartphone dari YouTube dan mengatur batas jumlah komentar yang akan diambil, sebelum menekan tombol untuk memulai proses analisis. Untuk lebih jelasnya dapat dilihat pada Gambar 3 di bawah ini.



Gambar 3. Tampilan Beranda

b. Tampilan Hasil Analisis

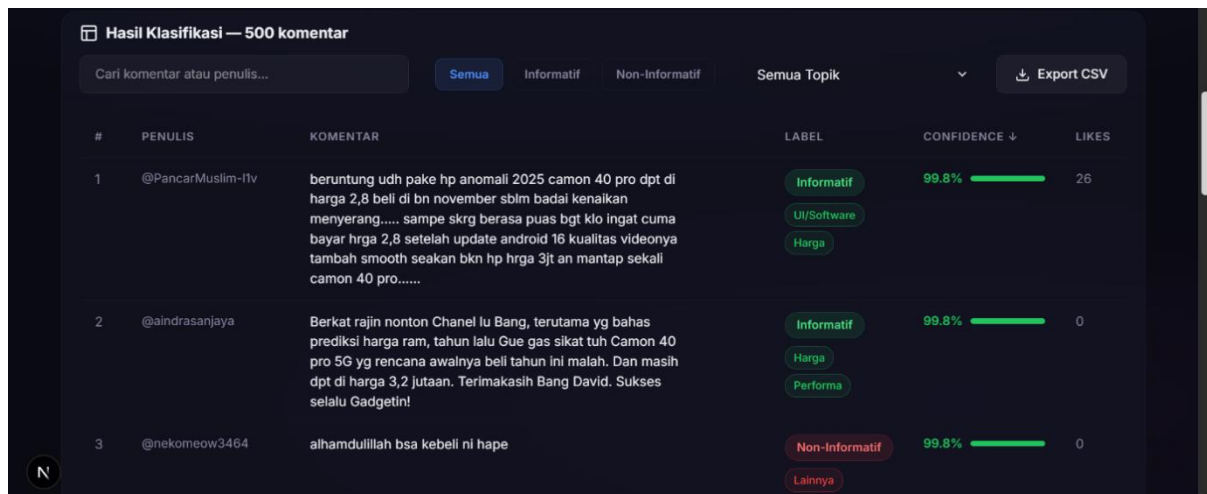
Berfungsi menampilkan ringkasan statistik hasil klasifikasi secara keseluruhan, baik dalam bentuk angka maupun visualisasi grafik, sehingga pengguna dapat langsung melihat proporsi komentar pada tiap kelas tanpa perlu membaca satu per satu. Ringkasan yang ditampilkan mencakup total komentar yang berhasil dianalisis serta jumlah komentar pada masing-masing kelas, yaitu Informatif dan Non-Informatif. Tampilan halaman ini dapat dilihat pada Gambar 4.



Gambar 4. Tampilan Hasil Analisis

c. Tampilan Hasil Klasifikasi

Berfungsi menampilkan hasil klasifikasi secara rinci per komentar, lengkap dengan label, tingkat keyakinan model, dan kategori topik, serta menyediakan fitur pencarian, penyaringan, dan ekspor data untuk menyimpan data komentar ke perangkat. Setiap baris data menampilkan nama penulis komentar, isi komentar, label hasil klasifikasi, skor tingkat keyakinan model dalam memberikan label tersebut, serta jumlah interaksi (likes) pada komentar. Rincian tampilan ini dapat diamati pada Gambar 5.

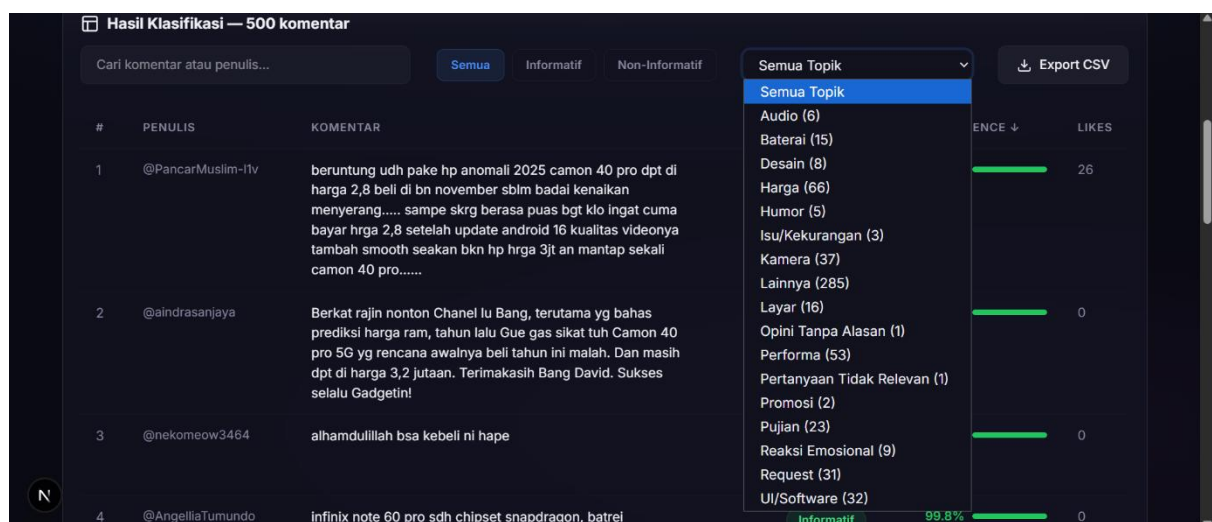


#	PENULIS	KOMENTAR	LABEL	CONFIDENCE ↓	LIKES
1	@PancarMuslim-Itv	beruntung udh pake hp anomali 2025 camon 40 pro dpt di harga 2,8 beli di bn november sbim badai kenaikan menyerang..... sampe skrg berasa puas bgt klo ingat cuma bayar hrge 2,8 setelah update android 16 kualitas videonya tambah smooth seakan bkn hp hrge 3jt an mantap sekali camon 40 pro.....	Informatif UI/Software Harga	99.8%	26
2	@alindrasanjaya	Berkat rajin nonton Chanel lu Bang, terutama yg bahas prediksi harga ram, tahun lalu Gue gas sikat tuh Camon 40 pro 5G yg rencana awalnya beli tahun ini malah. Dan masih dpt di harga 3,2 jutaan. Terimakasih Bang David. Sukses selalu Gadgetin!	Informatif Harga Performa	99.8%	0
3	@nekomeow3464	alhamdulillah bsa kebeli ni hape	Non-Informatif Lainnya	99.8%	0

Gambar 5. Tampilan Hasil Klasifiikasi

d. Tampilan Distribusi Sub-kategori

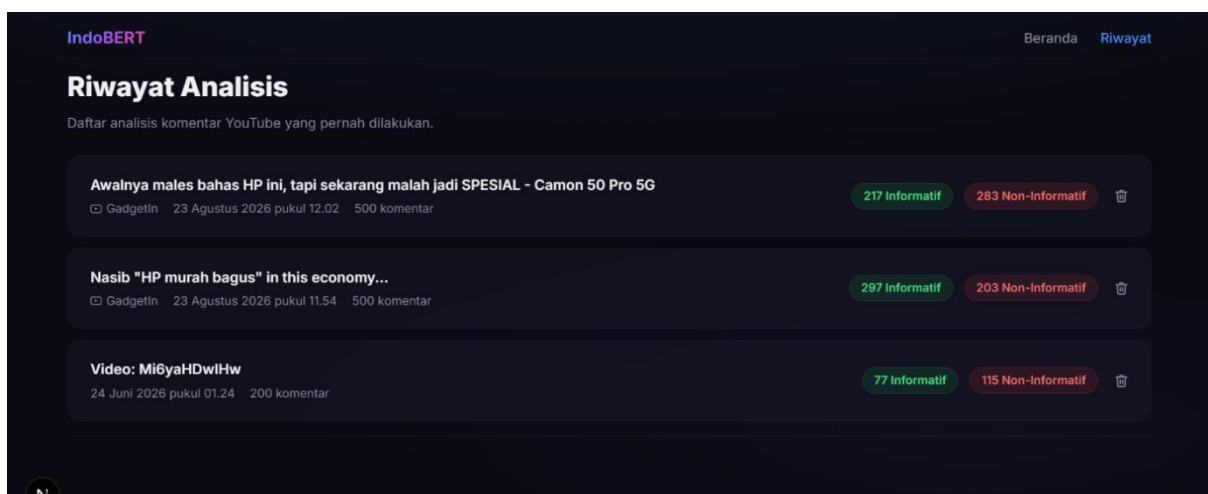
Berfungsi untuk memfilter komentar berdasarkan sub-kategori yang ingin ditampilkan pada halaman hasil klasifikasi. Fitur ini memungkinkan pengguna memilih aspek smartphone tertentu, seperti harga, kamera, baterai, performa, desain, layar, dan aspek lainnya sebagaimana telah dijelaskan pada Bab 2.6. Pengguna dapat memilih sub-kategori yang diinginkan untuk melihat distribusi komentar pada masing-masing kelas label, yaitu Informatif dan Non-Informatif. Dengan demikian, pengguna dapat lebih mudah mengetahui jumlah dan sebaran komentar berdasarkan topik yang dipilih serta melakukan analisis terhadap aspek smartphone yang dibahas dalam komentar. Gambar 6 berikut menunjukkan tampilan tersebut secara lebih rinci.



Gambar 6. Tampilan Distribusi Sub-kategori

e. Tampilan Riwayat Analisis

Berfungsi menyimpan dan menampilkan catatan analisis yang pernah dilakukan sebelumnya, sehingga pengguna dapat melihat kembali ringkasan hasil tanpa perlu mengulang proses analisis dari awal. Setiap riwayat menampilkan informasi video yang dianalisis, waktu pelaksanaan analisis, jumlah komentar yang diproses, serta ringkasan hasil klasifikasi pada setiap sesi. Tampilan riwayat analisis tersebut disajikan pada Gambar 7.



Gambar 7. Tampilan Riwayat Analisis

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, model IndoBERT mampu digunakan untuk mengklasifikasikan komentar berbahasa Indonesia pada video review smartphone di YouTube ke dalam kelas Informatif dan Non-Informatif dengan tingkat performa yang baik. Hal ini ditunjukkan dari hasil evaluasi model yang memperoleh akurasi sebesar 84,33%, presisi 83,66%, recall 85,33%, dan F1-score 84,49%.

Secara praktis, nilai akurasi sebesar 84,33% berarti dari setiap 100 komentar yang dianalisis, model mampu mengklasifikasikan sekitar 84 komentar dengan benar. Nilai presisi

sebesar 83,66% mengindikasikan bahwa sebagian besar komentar yang diprediksi sebagai Informatif memang benar-benar mengandung informasi yang relevan, sehingga pengguna tidak banyak disesatkan oleh hasil klasifikasi yang keliru. Sementara itu, nilai recall sebesar 85,33% menunjukkan bahwa model mampu menangkap sebagian besar komentar informatif yang sebenarnya ada dalam dataset, sehingga komentar yang berisi informasi penting tidak banyak terlewatkan. Nilai F1-score sebesar 84,49% yang merupakan rata-rata harmonis antara presisi dan recall mengkonfirmasi bahwa model memiliki keseimbangan yang baik dalam mendeteksi kedua kelas sekaligus.

Performa yang seimbang antara kelas Informatif (presisi 0,84, recall 0,85) dan kelas Non-Informatif (presisi 0,85, recall 0,83) mengindikasikan bahwa model tidak condong (bias) terhadap salah satu kelas. Kondisi ini didukung oleh distribusi data latih yang seimbang, yaitu 750 komentar per kelas, sehingga model mendapatkan representasi yang proporsional dari kedua kelas selama proses fine-tuning. Tingkat performa ini memberikan keyakinan yang memadai bahwa sistem dapat digunakan secara praktis untuk: (1) membantu calon konsumen menyaring komentar yang benar-benar informatif tanpa harus membaca ratusan komentar secara manual; (2) membantu produsen atau brand smartphone dalam memantau opini pengguna yang substantif terkait produk mereka; dan (3) mendukung penelitian atau analisis pasar yang membutuhkan data opini konsumen yang terstruktur dan berkualitas dari platform YouTube.

Adanya selisih kecil antara nilai presisi dan recall pada masing-masing kelas mengindikasikan bahwa masih terdapat sejumlah komentar yang salah klasifikasi, khususnya pada komentar yang menggunakan bahasa sangat informal, singkatan tidak umum, atau percampuran bahasa Indonesia dan bahasa Inggris (*code-mixing*) yang tidak terwakili dengan baik dalam data latih.

Selain itu, sistem yang telah dikembangkan juga berhasil diimplementasikan dalam bentuk aplikasi web yang mampu mengekstraksi komentar secara otomatis melalui YouTube Data API v3, mengklasifikasikannya menggunakan model IndoBERT, mengelompokkan hasil klasifikasi ke dalam sub-kategori berbasis kata kunci, serta menampilkan visualisasi statistik hasil analisis secara interaktif. Hasil Black Box Testing menunjukkan seluruh fungsi sistem berjalan sesuai spesifikasi, dan Usability Testing mengkonfirmasi bahwa antarmuka sistem mudah dipahami dan dioperasikan oleh pengguna yang tidak berlatar belakang teknis.

Hasil penelitian ini menunjukkan bahwa penerapan model IndoBERT berpotensi membantu berbagai pihak dalam menganalisis dan menyaring informasi dari kolom komentar YouTube secara lebih efektif, efisien, dan berbasis data. Klasifikasi komentar yang akurat dapat dimanfaatkan untuk mendukung pengambilan keputusan pembelian yang lebih terinformasi, pemantauan reputasi produk secara real-time, serta analisis sentimen dan tren opini konsumen terhadap smartphone tertentu. Untuk pengembangan selanjutnya, penelitian dapat difokuskan pada perluasan dataset dengan variasi video dan merek smartphone yang lebih beragam, pengembangan metode pengelompokan sub-kategori yang lebih robust terhadap variasi bahasa informal, serta eksplorasi model bahasa yang lebih mutakhir seperti IndoBERT versi yang lebih baru atau model berbasis arsitektur yang lebih besar untuk meningkatkan performa klasifikasi secara keseluruhan.

Daftar Pustaka

- Alvin, F., Anisa, N., & Winarsih, S. (2025). *Perbandingan Kinerja Model IndoBERT , IndoBERTweet , dan Algoritma Klasik pada Analisis Sentimen Isu Indonesia Gelap*. 7(3), 1601–1613. <https://doi.org/10.47065/bits.v7i3.8636>
- Cahyani, A. S., Fauzan, D. A., Padmavati, K., & Ekohandito, R. S. (2025). Pengaruh Konten Review Channel Youtube Gadgetin Terhadap Minat Beli Smartphone di Kalangan Gen Z. *Journal of Scientific Communication*, 7(1), 1–10.
- Kamilah, F. I., Rochman, M. Z., & Rosyid, H. Al. (2025). Analisis Sentimen Komentar Youtube Terhadap Perilisan Iphone 17 Series Menggunakan Algoritma Naive Bayes. *Journal of Information System, Informatics and Computing*, 9(2), 418–429. <https://doi.org/10.52362/jisicom.v9i2.2198>
- Karim, H. F., & Wibowo, A. P. (2025). Kinerja Metode Fine-Tuning IndoBERT untuk Klasifikasi Emosi Multi-Kelas pada Teks Informal Bahasa Indonesia. *Bulletin of Computer Science Research*, 6(1), 63–74. <https://doi.org/10.47065/bulletincsr.v6i1.850>
- Mahdi Faza, M. S., & Moch. Taufik. (2025). Penerapan Model Indobert Untuk Deteksi Potensi Sumber Stres Dalam Teks Media Sosial. *Jurnal Rekayasa Sistem Informasi Dan Teknologi*, 3(1), 271–283. <https://doi.org/10.70248/jrsit.v3i1.3046>
- Mefdi, R., & Halim, N. A. (2025). Analisis Penontonan Konten Ulasan Smartphone di Kanal YouTube GadgetIn dan Dampaknya terhadap Keputusan Pembelian Subscriber. *JURKOM : Jurnal Riset Komunikasi*, 8(2), 293–307.
- Monica, M., & Purwanto, A. (2026). *Comparative Performance Evaluation of Naïve Bayes and Logistic Regression for Indonesian YouTube Comment Sentiment Classification*. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 6(2), 934–943. <https://doi.org/10.57152/malcom.v6i2.2639>. 6(April), 934–943.
- Mulyono, H., Hartanti, R., & Rolando, B. (2024). Suara Konsumen Di Era Digital: Bagaimana Review Online Membentuk Perilaku Konsumen Digital. *Jurnal Bisnis Digital Dan Ekonomi Kreatif*, 1(1), 1–20. <https://journal.dinamikapublika.id/index.php/JUMDER>
- Nurjoko, & Rahardi, A. (2024). Model Indo-BERT Untuk Identifikasi Sentimen Kekerasan Verbal Di Twitter. *Jurnal Teknik*, 18(2), 583–593. <https://jurnal.polsri.ac.id/index.php/teknika/article/view/8897>
- Nuryadi, D., Metandi, F., Alam Hadiwijaya, N., Zainul Rohman, M., Hartanto, S., Syafrizal, A., & Yadie, E. (2025). Fine Tuning Indobert Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Tiket.Com Di Google Play Store. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 3577–3583. <https://doi.org/10.36040/jati.v9i2.13204>
- Pranata, J., Agustian, S., Jasril, J., & Haerani, E. (2024). Penggunaan Model Bahasa indoBERT pada metode Random Forest untuk Klasifikasi Sentimen dengan Dataset Terbatas. *Building of Informatics, Technology and Science (BITS)*, 6(3), 1668–1676. <https://doi.org/10.47065/bits.v6i3.6335>
- Salsabilla, T. R., & Nunik Pratiwi. (2024). Penerapan Support Vector Machine Untuk Analisis Sentimen pada X (Twitter) Mengenai Obat Penyebab Gagal Ginjal Akut pada Anak. *Jurnal Teknik Informatika Dan Komputer*, 3(2), 67–74. <https://doi.org/10.22236/jutikom.v3i2.16892>
- Simanjuntak, A., Lumbantoruan, R., Sianipar, K., Gultom, R., Simaremare, M., Situmeang, S., & Penggabean, E. (2024). Studi dan Analisis Hyperparameter Tuning IndoBERT

Dalam Pendeteksian Berita Palsu. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 13(1), 60–67.

Simarmata, A. A. P., & Sasongko, T. B. (2025). Sentiment Analysis on BRIImo Application Reviews Using IndoBERT. *Journal of Applied Informatics and Computing*, 9(3), 851–862. <https://doi.org/10.30871/jaic.v9i3.8162>

Sumartha, D. K. (2025). Implementasi IndoBERT untuk Analisis Sentimen Opini Publik terhadap Kebijakan Kenaikan UKT di Era Pemerintahan. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3S1). <https://doi.org/10.23960/jitet.v13i3S1.7880>.

CC BY-SA 4.0 (Attribution-ShareAlike 4.0 International).

This license allows users to share and adapt an article, even commercially, as long as appropriate credit is given and the distribution of derivative works is under the same license as the original. That is, this license lets others copy, distribute, modify and reproduce the Article, provided the original source and Authors are credited under the same license as the original.

