



Rancang Bangun Sistem Analisis Sentimen Opini Publik Multi-Source Berbasis Pemahaman Konteks Menggunakan *Fine-Tuning* IndoBERT (Studi Kasus Universitas Negeri Padang)

Reyhan Putra Febrian¹, Yulia Fatmi², Syafrijon³, Randi Proska Sandra⁴

¹Program Studi Informatika, Fakultas Teknik, Universitas Negeri Padang, Indonesia

²Universitas Negeri Padang, Indonesia

Correspondence: E-mail: reyhanputrafebrian3@gmail.com

ABSTRACT

Penelitian ini bertujuan untuk mengembangkan sistem analisis sentimen opini publik *multi-source* berbasis aplikasi web interaktif guna memfasilitasi pemantauan persepsi masyarakat terhadap Universitas Negeri Padang. Pendekatan yang digunakan mengintegrasikan model pemrosesan bahasa alami berbasis *Transformer*, yakni *fine-tuned IndoBERT*, yang dieksekusi melalui arsitektur Client Server (*Next.js* dan *FastAPI*), serta dilengkapi dengan mekanisme *Hybrid Inference Logic (Lexicon-based override)* untuk memproses data secara *batch processing*. Hasil evaluasi menggunakan *Confusion Matrix* menunjukkan bahwa model IndoBERT mencapai akurasi sebesar 95% dalam mengklasifikasikan 3.411 baris data uji, dengan *precision* tertinggi mencapai 97% pada sentimen positif dan *recall* sebesar 96% pada sentimen negatif. Implementasi infrastruktur ini terbukti andal dalam menangani beban komputasi teks dari berbagai platform digital (X, Instagram, Portal Berita) tanpa kendala latensi. Dampak dari inovasi ini adalah terciptanya solusi *Decision Support System (DSS)* yang skalabel, akurat, dan adaptif untuk memberikan landasan evaluasi dan pengambilan keputusan institusional yang objektif.

ARTICLE INFO

Article History:

Submitted/Received 20 July 2026

First Revised 21 July 2026

Accepted 23 July 2026

First Available online 24 July 2026

Publication Date 24 July 2026

Keyword:

Analisis Sentimen, Decision Support System, IndoBERT, Inferensi Hibrida, Multi-Source, NLP, Opini Publik.

1. INTRODUCTION

Perkembangan pesat teknologi informasi dan komunikasi di era digitalisasi telah memicu transformasi masif dalam cara masyarakat mengekspresikan pendapat, kritik, serta evaluasi terhadap berbagai institusi layanan publik. Ekosistem media sosial dan platform digital seperti X (sebelumnya Twitter), Instagram, forum daring, hingga kolom komentar portal berita online kini telah *bermetamorfosis* menjadi ruang publik virtual (*virtual public sphere*) yang sangat aktif dan dinamis dalam membentuk opini komunal (Geni et al., 2023). Melalui platform-platform tersebut, pengguna internet memiliki kebebasan untuk berpartisipasi secara langsung dalam penciptaan serta pertukaran konten (*user generated content*), yang memungkinkan sebuah opini menyebar secara eksponensial dalam skala yang luas. Tumpukan data opini publik yang berkembang di ruang digital ini pada umumnya berwujud data teks tidak terstruktur (*unstructured text data*) dengan volume yang masif (*big data*). Meskipun tidak terstruktur, data teks tersebut menyimpan nilai strategis yang sangat tinggi, mengingat di dalamnya terkandung representasi nyata mengenai persepsi, sikap emosional, dan penilaian masyarakat terhadap suatu kebijakan, layanan, maupun citra suatu institusi (Nugraha, 2020).

Dalam konteks tata kelola pendidikan tinggi, opini publik berperan sebagai instrumen indikator performa yang krusial untuk mengevaluasi kualitas layanan akademik, efektivitas kebijakan institusi, serta reputasi perguruan tinggi di mata masyarakat luas. Universitas Negeri Padang (UNP), sebagai salah satu perguruan tinggi negeri terkemuka, memiliki tingkat interaksi publik yang sangat intensif di ruang digital. Berbagai pemangku kepentingan mulai dari mahasiswa aktif, calon mahasiswa, alumni, hingga masyarakat umum secara konsisten dan proaktif menyampaikan opini mereka terkait UNP. Wacana yang beredar meliputi berbagai topik krusial, seperti transparansi Uang Kuliah Tunggal (UKT), kualitas fasilitas kampus, hingga prestasi akademik dan non-akademik civitas akademika (Yoga Pratama et al., 2025). Opini-opini tersebut tersebar melintasi berbagai sumber platform (*multi source*) dan memiliki karakteristik *linguistik* yang sangat *heterogen*. Kondisi ini menyebabkan proses evaluasi sentimen secara konvensional atau manual menjadi sangat tidak rasional, memakan waktu yang ekstensif, dan sangat rentan terdistorsi oleh bias subjektivitas pengamat manusia.

Untuk mengatasi kompleksitas tersebut, penerapan Analisis Sentimen (*Sentiment Analysis*) atau penambangan opini (*opinion mining*) yang bernaung di bawah disiplin *Natural Language Processing* (NLP) menjadi sebuah solusi komputasional yang mutlak diperlukan. Analisis sentimen memungkinkan *ekstraksi* dan *klasifikasi* polaritas emosi di dalam suatu entitas tekstual ke dalam kategori diskrit, seperti sentimen positif, netral, maupun negatif, secara otomatis dan sistematis (Rizkia et al., 2025). Kendati demikian, implementasi NLP untuk memproses dialek digital berbahasa Indonesia dihadapkan pada rintangan struktural yang signifikan. Algoritma *machine learning* konvensional seperti *Support Vector Machine* (SVM) atau *Naive Bayes* sering kali mengalami kegagalan *kognitif* (*cognitive breakdown*) ketika dihadapkan pada teks media sosial yang sarat dengan bahasa gaul (*slang*), akronim spontan, dan penyimpangan kaidah tata bahasa baku. Rintangan ini menjadi semakin kompleks dengan merebaknya fenomena majas sarkasme atau ironi di media sosial, di mana netizen kerap membalut kritik pedas atau kekecewaan mendalam menggunakan diksi yang secara harfiah bersentimen positif. Ketiadaan struktur sintaksis baku serta dominasi makna implisit pada kalimat sindiran ini meruntuhkan keandalan metode

berbasis leksikon tradisional, karena mesin gagal menangkap intensi atau niat asli dari sang penulis (Munir, 2025).

Terobosan fundamental dalam mengatasi ambiguitas semantik tersebut hadir melalui arsitektur komputasi *deep learning* berbasis *Transformer*, khususnya varian *Bidirectional Encoder Representations from Transformers* (BERT) (Koto et al., 2020). Berbeda dengan arsitektur sekuensial terdahulu, *BERT* memanfaatkan mekanisme atensi mandiri (*self-attention*) dua arah (*bidirectional*) yang memungkinkannya untuk membaca dan mengevaluasi bobot keterkaitan seluruh kata di dalam sebuah kalimat secara serempak. Kemampuan ini memberikan BERT kepekaan konteks yang luar biasa tajam. *IndoBERT*, sebagai inkarnasi *BERT* yang dilatih menggunakan pangkalan korpus linguistik *Indo4B* (berisi lebih dari 4 miliar token kata berbahasa Indonesia), telah terbukti sangat superior dalam menangani karakteristik linguistik lokal dan memberikan performa *klasifikasi* sentimen yang presisi. Melalui rekayasa *fine tuning*, parameter bobot pada *IndoBERT* dapat dikalibrasi ulang menggunakan dataset spesifik domain institusi untuk menciptakan mesin *inferensi* kecerdasan buatan yang sangat akurat (Maysara & Muljono, 2026).

Meskipun arsitektur *IndoBERT* telah diakui keunggulannya dalam berbagai studi komparatif, masih terdapat kesenjangan riset (*research gap*) yang nyata dalam hal implementasi praktisnya sebagai sistem pendukung keputusan (*Decision Support System*) yang utuh. Mayoritas riset terdahulu cenderung terjebak pada eksperimen laboratorium berbasis data tunggal (*single source dataset*) dan belum diwujudkan ke dalam bentuk perangkat lunak siap pakai. Selain itu, belum tersedia platform berbasis web terintegrasi yang dirancang secara khusus untuk menghimpun, memproses (*batch processing*), dan memvisualisasikan opini publik terkait Universitas Negeri Padang dari berbagai sumber digital sekaligus (*multi source*). Oleh karena itu, penelitian ini bertujuan untuk merancang dan membangun sistem analisis sentimen opini publik berbasis arsitektur *client server* modern yang ditenagai oleh model *fine tuning* *IndoBERT* serta mekanisme logika inferensi hibrida (*Hybrid Inference Logic*). Inovasi ini diharapkan mampu menghadirkan produk teknologi informasi yang andal, skalabel, dan aplikatif dalam menjembatani kebutuhan evaluasi institusional secara *komprehensif*, *objektif*, dan berbasis data (*data driven decision making*).

2. METHODS

Penelitian ini direalisasikan dengan mengadopsi metodologi rekayasa perangkat lunak Software Development Life Cycle (SDLC) menggunakan model Waterfall. Pemilihan model sekuensial ini didasarkan pada pertimbangan akademis dan teknis bahwa seluruh spesifikasi fungsional (system requirements) untuk platform analisis sentimen ini telah terdefinisi secara komprehensif, statis, dan tidak menuntut perubahan radikal di tengah siklus pengembangan. Karakteristik Waterfall yang mewajibkan penyelesaian tuntas dan validasi absolut pada setiap fasemulai dari analisis kebutuhan, desain arsitektur, implementasi kode, hingga pengujian sistem memberikan jaminan stabilitas yang krusial bagi kelancaran alur kerja (pipeline) machine learning. Dalam pengolahan data teks berskala besar, tahapan prapemrosesan (preprocessing) hingga klasifikasi inferensi memiliki tingkat keterikatan sekuensial yang tidak dapat dipisahkan (Yanti et al., 2024). Secara terstruktur, metodologi penelitian ini diklasifikasikan ke dalam empat fase utama operasional AI, yakni: akuisisi data multi-source, prapemrosesan teks, fine-tuning model *IndoBERT*, serta integrasi sistem dan perancangan antarmuka.

2.1 Data Acquisition and Multi-Source Integration

Kualitas dan varians data merupakan fondasi vital bagi performa model kecerdasan buatan. Untuk menghindari bias demografis dan merepresentasikan opini publik secara utuh, penelitian ini mengimplementasikan strategi akuisisi data *Multi Source*. Dataset dihimpun dari berbagai *episentrum* perbincangan digital, mencakup platform X (Twitter), kolom komentar Instagram pada akun-akun afiliasi kampus, serta forum diskusi portal berita daring. Proses ekstraksi data dilakukan menggunakan teknik web *scraping* terarah dengan menargetkan kata kunci spesifik (keywords) seperti "UNP Padang", "UKT Universitas Negeri Padang", dan "Fasilitas UNP". Kumpulan teks opini mentah (raw text) ini kemudian diagregasi dan dikonsolidasikan ke dalam format *Comma Separated Values* (CSV). Dataset berbasis CSV inilah yang bertindak sebagai objek masukan utama (*input*) yang akan diunggah oleh pengguna (Admin Fakultas) melalui antarmuka dashboard untuk kemudian diproses secara terpusat (*batch processing*) oleh sistem *backend*.

2.2 Text Preprocessing Pipeline

Teks opini publik di media sosial dikenal dengan karakteristiknya yang sangat tidak terstruktur (*noisy text*). Memasukkan data mentah secara langsung ke dalam model *klasifikasi* akan memicu *degradasi* akurasi yang fatal. Oleh karena itu, sistem ini dilengkapi dengan modul *Text Preprocessing otomatis* yang bertugas membersihkan dan menormalisasi setiap baris data sebelum masuk ke fase inferensi (Khairul & Perdana, 2024). Modul prapemrosesan ini menjalankan serangkaian fungsi *sekuensial*, dimulai dengan *Cleansing berbasis ekspresi reguler (Regex)* untuk mengeliminasi elemen derau seperti tautan (*URL/ Hyperlink*), tagar (*hashtag*), penyebutan akun (*@mention*), indikator retweet (*RT*), karakter numerik, hingga tanda baca dan simbol *non-alfanumerik*. Setelah teks bersih dari *noise*, tahapan dilanjutkan dengan *Tokenisasi*. Pada fase ini, sistem menginisialisasi tokenizer bawaan dari IndoBERT (*AutoTokenizer*). Tokenizer ini memecah kalimat menjadi unit sub-kata (subwords) yang kompatibel dengan vokabulari model. Untuk menjaga stabilitas dimensi matriks pada arsitektur *neural network*, sistem menerapkan parameter batas maksimal panjang (*max_length*) sebesar 256 token. Teknik pemotongan (*truncation*) diaplikasikan secara paksa pada opini yang melampaui batas tersebut, sementara teknik penambahan padding (token [PAD]) diinjeksikan pada kalimat pendek guna menghasilkan struktur input tensor berdimensi seragam.

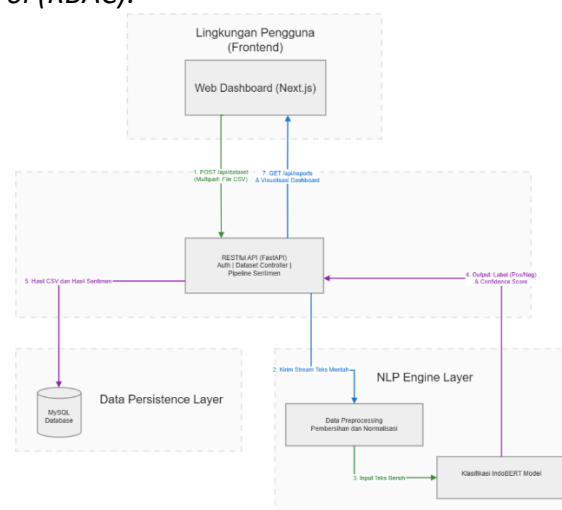
2.3 Model Fine-Tuning and Hybrid Inference Logic

Inti kecerdasan buatan pada sistem ini bersumber pada model IndoBERT base (Koto et al., 2020) yang telah dioptimasi melalui mekanisme *Fine Tuning*. Proses kalibrasi bobot parameter ini dieksekusi secara *intensif* pada lingkungan komputasi awan berakselerasi GPU. Dataset yang telah dinormalisasi dipecah ke dalam tiga subset dengan rasio 60:20:20, masing-masing dialokasikan sebagai data latih (*training*), data validasi (*validation*), dan data uji (*testing*). Proses fine-tuning dikonfigurasi dengan menyetel parameter laju pembelajaran (*learning rate*) pada angka konservatif 2×10^{-5} dan ukuran tumpukan (*batch size*) sebanyak 16 sampel per iterasi. Untuk memastikan model mencapai titik konvergensi analitis maksimal dalam memahami konteks majas sarkasme dan sentimen lokal, pelatihan dilangsungkan selama 30 siklus (epochs). Model yang telah terlatih matang ini kemudian

diekspor dalam format PyTorch untuk diintegrasikan secara lokal pada peladen (*server*). Sebagai bentuk mitigasi terhadap anomali prediksi dan memastikan keandalan sistem dalam berbagai skenario ekstrem, sistem ini mengimplementasikan algoritma *Adaptive Hybrid Inference Logic*. Logika ini menggabungkan kecanggihan pemahaman kontekstual IndoBERT dengan ketegasan leksikal dari kamus *Lexicon konvensional*. Saat teks diproses, IndoBERT menghasilkan prediksi kelas sentimen beserta skor probabilitas (*Confidence Score*) melalui fungsi aktivasi *Softmax*. Secara paralel, fungsi analitik *Lexicon* mengkalkulasi bobot polaritas kata per kata (*Lexicon Score*). Algoritma Hibrida beroperasi dengan tiga aturan keputusan utama (*Decision Gates*): (1) Jika IndoBERT memiliki *confidence score* tinggi ($\geq 60\%$), keputusan mutlak berada pada model *deep learning*, (2) *Fallback Trigger*: Jika *confidence score* IndoBERT rendah ($< 60\%$) namun *Lexicon* menunjukkan sentimen yang kuat (skor ≥ 2), sistem akan mengambil alih keputusan menggunakan label *Lexicon* guna menghindari *ambiguitas*; (3) *Override Trigger*: Jika skor *Lexicon* sangat dominan (skor absolut ≥ 4), sistem secara paksa menganulir prediksi IndoBERT dan menetapkan sentimen berdasarkan bobot *Lexicon*. Mekanisme ini menjamin bahwa sistem tetap tangguh meskipun dihadapkan pada teks dengan sintaksis yang sangat kacau.

2.4 System Architecture and Database Schema

Secara arsitektural, platform ini dirancang dengan memisahkan wilayah antarmuka pengguna (*Frontend*) dan logika bisnis serta komputasi AI (*Backend*). Lingkungan *Frontend* dibangun memanfaatkan kerangka kerja *Next.js* yang dikombinasikan dengan *Tailwind CSS* untuk menghadirkan antarmuka *Single Page Application (SPA)* yang sangat responsif, intuitif, dan kaya akan visualisasi interaktif (Muhammad Samsu Hidayattullah et al., 2025). Di sisi lain, layanan *Backend* digerakkan oleh *framework FastAPI* berbasis *Python* yang secara natif sangat optimal dalam menangani beban komputasi *asynchronous* dan injeksi model *machine learning*. Fasilitas penyimpanan persisten (*Data Persistence*) dikelola oleh sistem manajemen basis data relasional *MySQL*. Skema database dinormalisasi ke dalam relasi tabel entitas yang kohesif, meliputi tabel pengguna (*users*), dataset mentah (*datasets*, *opinions*), hasil prapemrosesan (*preprocessed_texts*), hasil klasifikasi sentimen (*analysis_results*), dan agregasi pelaporan (*reports*) (Rafianto & Voutama, 2025). Interkoneksi antara frontend, backend, dan basis data ini dilindungi menggunakan protokol enkripsi *JSON Web Token (JWT)* untuk mengamankan manajemen sesi berdasarkan struktur *Role-Based Access Control (RBAC)*.



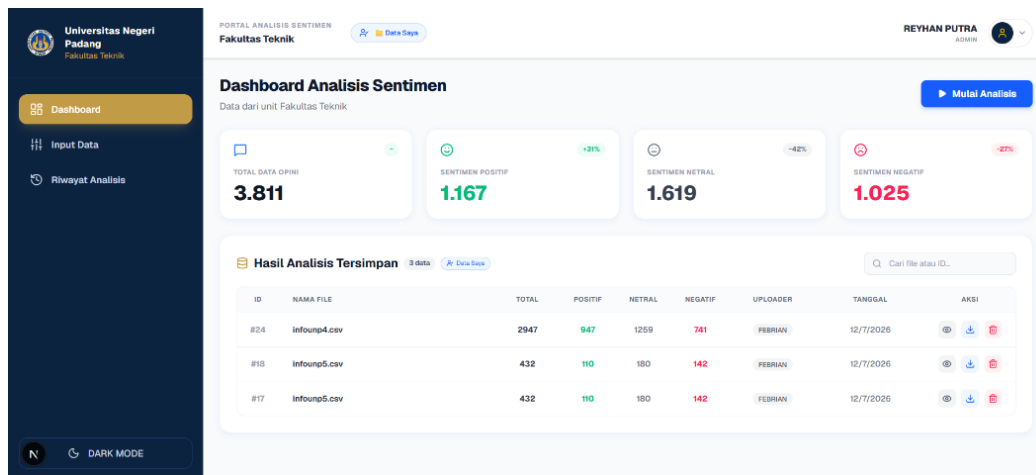
Gambar 1. Topologi Arsitektur Sistem Client-Server

3. RESULTS AND DISCUSSION

Bagian ini mendemonstrasikan hasil perwujudan sistem dari segi antarmuka perangkat lunak, evaluasi metrik kinerja algoritma *deep learning*, serta uji fungsionalitas sistem secara holistik dalam menangani beban pemrosesan sentimen opini publik berbasis batch.

3.1 System Interface and Functional Implementation

Platform analisis sentimen berhasil direalisasikan ke dalam wujud web dashboard interaktif yang sepenuhnya operasional. Tata letak sistem mematuhi prinsip desain User-Centered Design, memastikan pengguna institusional (seperti pimpinan unit atau Admin Fakultas) dapat menavigasi fungsi analitik yang kompleks tanpa memerlukan latar belakang teknis komputasi. Setelah melewati portal autentikasi (Login) yang divalidasi oleh token JWT, Super Admin diberikan akses penuh ke halaman Dashboard Global. Halaman ini bertindak sebagai pusat komando (command center) yang merender ringkasan komprehensif dari seluruh aktivitas sentimen di lingkungan Universitas Negeri Padang.

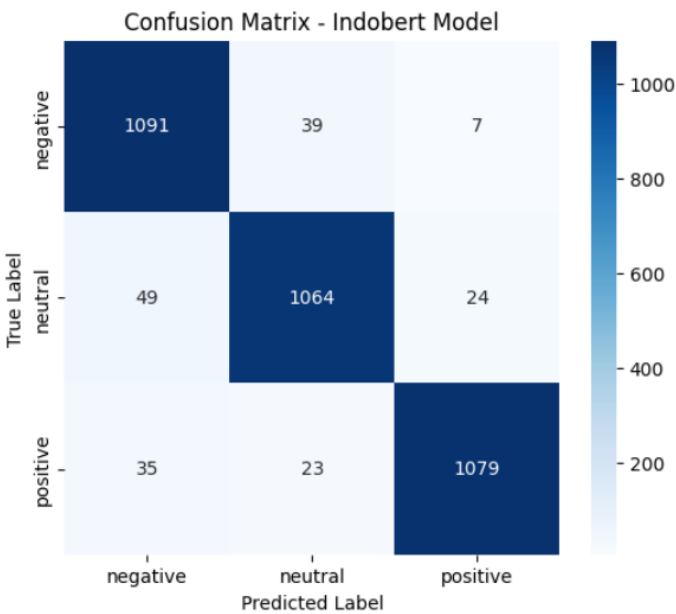


Gambar 2. Implementasi Antarmuka Dashboard Global Sentimen

Gambar 2 memperlihatkan visualisasi data analitik yang dirender secara real-time dari basis data MySQL untuk unit kerja spesifik (seperti Fakultas Teknik). Komponen Key Performance Indicator (KPI) pada bagian atas dashboard menampilkan metrik vital seperti Total Data Opini (3.811), Sentimen Positif (1.167), Sentimen Netral (1.619), dan Sentimen Negatif (1.025). Di bawah kartu ringkasan KPI, sistem dilengkapi dengan tabel "Hasil Analisis Tersimpan" yang memuat daftar riwayat unggahan berkas dataset (seperti infounp4.csv dan infounp5.csv) lengkap dengan informasi ID, Nama File, Total Data, rincian jumlah sentimen, Uploader, Tanggal, serta kolom aksi interaktif (ikon lihat, unduh, dan hapus). Di sisi Admin Fakultas, sistem menyediakan halaman Input Data khusus untuk mengakomodasi pengunggahan berkas dataset CSV. Modul ini dilengkapi dengan validasi format dan batasan ukuran berkas untuk menjamin stabilitas pemrosesan backend. Segera setelah CSV diunggah, mesin FastAPI akan memecah baris data, menjalankan cleansing, mengeksekusi model IndoBERT, dan memperbarui rekapitulasi data di dashboard secara seketika (seamless).

3.4 Model Accuracy and Performance Evaluation

Kemanjuran sistem sebagai perangkat pengambil keputusan sangat bergantung pada ketepatan mesin NLP dalam mengklasifikasikan sentimen. Kinerja model Fine-Tuned IndoBERT dievaluasi menggunakan instrumen Confusion Matrix terhadap subset data uji (testing data) yang berisikan 3.411 baris opini publik aktual yang belum pernah dilatihkan kepada model. Evaluasi kuantitatif ini dihitung menggunakan metrik Accuracy, Precision (tingkat kepresisian tebakan), Recall (sensitivitas penjarangan data), dan F1-Score (rata-rata harmonis) untuk menyeimbangkan penilaian pada kelas distribusi sentimen yang fluktuatif.



Gambar 3. Matriks Kebingungan (Confusion Matrix) Model IndoBERT

Berdasarkan simulasi Confusion Matrix (Gambar 3), dominasi nilai pada diagonal utama membuktikan bahwa model jarang melakukan misklasifikasi. Rekapitulasi komprehensif dari pencapaian performa model pada masing-masing kelas sentimen didokumentasikan dalam Tabel 1.

Kelas Sentimen	Precision	Recall	F1-Score	Support (Total Data)
Negatif	0.93	0.96	0.94	1137
Netral	0.94	0.94	0.94	1137
Positif	0.97	0.95	0.96	1137
Akurasi Keseluruhan (Overall Accuracy)			0.95	3411

Tabel 1. Laporan Metrik Performa Klasifikasi Sentimen Akhir (Final Test Set Report)

Merujuk pada Tabel 1, model deep learning mencatatkan performa yang sangat impresif dengan pencapaian akurasi keseluruhan (Overall Accuracy) sebesar 95%. Secara

lebih mendalam, evaluasi per kelas memberikan wawasan yang esensial mengenai ketajaman model. Pada pengujian sentimen Negatif, metrik Recall berhasil menembus angka 96%. Tingginya sensitivitas ini mengonfirmasi kapabilitas model dalam menjangkau nyaris keseluruhan opini bermuatan kritik tajam, keluhan pelayanan, hingga ketidakpuasan struktural dari mahasiswa tanpa banyak celah keterlewatan (False Negative sangat minim). Kapasitas ini sangat krusial dalam domain layanan publik karena memungkinkan institusi untuk mendeteksi krisis lebih awal (early warning system).

Di spektrum yang berlawanan, kelas sentimen Positif mencetak rekor Precision tertinggi di angka 97%, didampingi nilai Recall sebesar 95%. Pencapaian kepresisian yang nyaris sempurna ini menunjukkan bahwa arsitektur Transformer sangat sukses mereduksi kesalahan identifikasi positif semu (False Positive). Keberhasilan ini tidak lepas dari peran arsitektur 12 lapisan atensi (attention heads) pada IndoBERT yang

mampu melacak jarak semantik antar-kata dengan sangat presisi, sehingga mampu memilah dengan tegas antara apresiasi tulus dari civitas akademika dengan frasa pujian hiperbolis yang sebenarnya mengandung majas sarkasme atau ironi (sindiran halus). Stabilitas pada kelas Netral (Precision 94%, Recall 94%) juga membuktikan bahwa model mampu mengklasifikasikan kalimat deklaratif atau informatif murni tanpa memaksakan kecenderungan polaritas secara artifisial.

3.5 System Functional Evaluation (Black Box Testing)

Guna memastikan kelayakan platform untuk dioperasikan dalam lingkungan produksi (production ready), pengujian fungsional komprehensif dilakukan menggunakan pendekatan Black Box Testing. Fokus pengujian diletakkan pada reliabilitas sistem dalam menangani alur kerja pemrosesan batch yang mengkonsumsi memori komputasi tinggi. Skenario uji mencakup: (1) Otorisasi RBAC pengguna, (2) Validasi format ekstensi saat proses unggah dataset CSV, (3) Eksekusi otomatis pipeline Text Preprocessing dan NLP inference pada ribuan baris data, (4) Sinkronisasi penyimpanan hasil inferensi ke dalam MySQL, serta (5) Kemampuan sistem dalam mengeksport dan mengunduh laporan akhir berformat CSV ke penyimpanan lokal perangkat klien.

Hasil pengujian mendemonstrasikan tingkat keberhasilan 100% pada seluruh skenario. Sistem terbukti sangat tangguh (robust); proses integrasi antara form unggahan Next.js dan controller FastAPI berjalan sinkron. Mesin tidak mengalami kegagalan memori (memory leak) ataupun crash saat ditugaskan untuk membersihkan, menokenisasi, dan mengevaluasi lebih dari 3.000 entitas komentar secara simultan di latar belakang. Proses pengunduhan laporan juga dieksekusi dengan sempurna melalui pembuatan URL objek lokal di peramban, memastikan data sensitif dapat direkapitulasi secara aman oleh pimpinan fakultas.

3.6 Discussion

Inisiasi platform analisis sentimen ini menandai pergeseran paradigma dari metode evaluasi umpan balik konvensional menuju tata kelola persepsi publik berbasis komputasi cerdas (smart governance) di lingkungan Universitas Negeri Padang. Integrasi arsitektur IndoBERT dengan mekanisme Hybrid Inference Logic (menggunakan Lexicon fallback) terbukti berhasil menetralkan kelemahan struktural pada algoritma machine learning

tradisional (seperti Naive Bayes atau SVM). Model pendahulu sering kali terdistorsi oleh tingginya tingkat redundansi dan ambiguitas slang pada bahasa netizen (Afifah & Dedi Gunawan, 2025; Daniati et al., 2026), sedangkan kerangka kerja Transformer dalam penelitian ini justru berkembang pesat berkat kemampuannya memodelkan relasi bidireksional pada setiap kalimat.

Pencapaian akurasi sebesar 95% secara telak melampaui standar target eksperimental dan membuktikan bahwa sistem ini lebih dari sekadar prototipe akademis. Dari perspektif manajerial, dasbor visualisasi (Dashboard Global) memberikan nilai utilitas komersial (EdTech and Business Value) yang tak terbantahkan. Pimpinan universitas kini memiliki instrumen pengukur barometer emosi massa yang objektif. Analitik tren opini bulanan dan komparasi distribusi sentimen lintas-fakultas memungkinkan jajaran rektorat untuk melakukan mitigasi dini terhadap isu-isu negatif yang tengah memanas, sembari memproyeksikan strategi optimalisasi layanan akademik di masa depan berdasarkan persepsi positif yang telah terbangun kokoh (data-driven strategy).

4. CONCLUSION

Penelitian ini telah sukses merancang, mengembangkan, dan memvalidasi sebuah sistem analisis sentimen opini publik multi-source berbasis aplikasi web interaktif yang ditenagai oleh kecerdasan buatan. Implementasi arsitektur fine-tuning IndoBERT terbukti sangat revolusioner dalam memecahkan kompleksitas pemahaman konteks, majas sarkasme, dan karakteristik bahasa slang Indonesia yang marak di platform digital. Keandalan modul NLP ini dibuktikan dengan pencapaian akurasi klasifikasi sebesar 95% dari 3.411 data uji, dengan capaian spesifik precision 97% untuk sentimen positif dan recall 96% untuk sentimen negatif. Penggabungan prediksi deep learning dengan logika leksikon hibrida memberikan garansi keamanan berlapis terhadap fluktuasi probabilitas data teks. Infrastruktur komputasi berbasis Client-Server (Next.js dan FastAPI) yang dirancang juga terbukti tangguh dalam mengeksekusi pemrosesan data bervolume tinggi secara batch tanpa mengalami kendala latensi maupun anomali memori. Secara praktis, inovasi ini memberikan sumbangsih besar berupa Decision Support System (DSS) yang skalabel dan objektif, yang memfasilitasi institusi pendidikan tinggi seperti Universitas Negeri Padang untuk melakukan pemantauan, evaluasi, dan perumusan kebijakan layanan berbasis data secara responsif, transparan, dan terukur di era disruptif digital.

5. ACKNOWLEDGMENT

Penulis menyampaikan apresiasi dan penghargaan yang setinggi-tingginya kepada Dosen Pembimbing yang telah memberikan arahan akademis, kritik konstruktif, serta bimbingan teknis yang sangat berharga selama penyusunan riset ini. Ucapan terima kasih yang tulus juga ditujukan kepada Departemen Elektronika, Fakultas Teknik, Universitas Negeri Padang atas dukungan moral serta fasilitasi ekosistem penelitian yang memadai, sehingga purwarupa sistem kecerdasan buatan ini dapat terealisasi dengan optimal.

6. AUTHORS' NOTE

Penulis dengan ini mendeklarasikan bahwa tidak terdapat konflik kepentingan (*conflict of interest*) baik secara intelektual maupun finansial dalam publikasi naskah ini. Artikel ini disusun secara murni dari hasil penelitian orisinal (*original research*) dan tidak melanggar kode etik plagiarisme. Seluruh sitasi dan rujukan terhadap literatur pihak ketiga telah dicantumkan secara transparan di dalam badan teks dan daftar pustaka.

7. REFERENCES

- Afifah, L. A., & Dedi Gunawan. (2025). Analisis Sentimen Pengguna X terhadap IKN Menggunakan Word2Vec dan IndoBERT Luthfiana. 7, 864–874. <https://doi.org/10.30865/json.v7i3.9468>
- Daniati, E., Nugroho, A., Ristyawan, A., & Utama, H. (2026). Perbandingan Kinerja Algoritma SVM, LSTM, dan Fine-tuned IndoBERT dalam Analisis Sentimen Opini Masyarakat Indonesia terhadap Mobil Listrik. 5(1), 1–13. <https://doi.org/10.59095/ijcsr.v5i1.245>
- Geni, L., Yulianti, E., & Sensuse, D. I. (2023). Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using Bert Language Models. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, 9(3), 746–757. <https://doi.org/10.26555/jiteki.v9i3.26490>
- Khairul, M. F. R., & Perdana, R. S. (2024). Arsitektur Sistem Percakapan Otomatis Berbahasa Indonesia dengan Normalisasi Bahasa Informal Menjadi Baku. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(5), 1009–1016. <https://doi.org/10.25126/jtiik.2024117984>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, 757–770. <https://doi.org/10.18653/v1/2020.coling-main.66>
- Maysara, A. L., & Muljono. (2026). Optimalisasi Akurasi dan Stabilitas Analisis Sentimen Ulasan E-Commerce Indonesia melalui Fine-Tuning Transformer IndoBERT. 17(01), 9–17. <https://doi.org/10.35970/infotekmesin.v17i1.3037>
- Muhammad Samsu Hidayattullah, Muhammad Izman Herdiansyah, Heri Suroyo, & Ade Putra. (2025). Pengembangan Sistem Pendaftaran dan Pengelolaan Laporan Magang di Kominfo Kota Palembang Berbasis Website Menggunakan Next Js dan Express. *Jurnal Informatika Dan Tekonologi Komputer (JITEK)*, 5(3), 90–99. <https://doi.org/10.55606/jitek.v5i3.8397>
- Munir, R. A. (2025). Analisis Sentimen Cuitan Di Media Sosial X Tentang Program Makan Bergizi Gratis Dengan Metode Nlp. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3), 589–595. <https://doi.org/10.23960/jitet.v13i3.6912>
- Nugraha, B. (2020). Metode Klasifikasi Analisis Sentimen pada Media Sosial. *Syntax : Jurnal Informatika*, 9(2), 119–123. <https://doi.org/10.35706/syji.v9i2.3593>
- Rafianto, G., & Voutama, A. (2025). Implementasi Basis Data Terstruktur Dengan Pencegahan Sql Injection Pada Sistem Manajemen Penjualan. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(2). <https://doi.org/10.23960/jitet.v13i2.6354>
- Rizkia, A. S., Wufron, & Roji, F. F. (2025). Sentiment Analysis of Coretax: A Comparison of Manual, Transformers- Based, and Lexicon-Based Data Labeling on IndoBERT Performance. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3), 1037–1048. <https://doi.org/10.57152/malcom.v5i3.2151>
- Yanti, N. M. N. F., Paramitha, A. A. I. I., & Dewi, E. G. A. (2024). Rancang Bangun Sistem

Informasi Bimbingan Konseling Pada Sma Negeri 1 Baturiti Dengan Metode Waterfall.
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI), 13(3), 117–
128. <https://doi.org/10.23887/karmapati.v13i3.87662>

Yoga Pratama, A., Ananda Sanjaya, G., Khairunisa Lubis, N., & Rangga Aditya, M. (2025).
Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada
Platform Media Sosial. *METIK JURNAL VOLUME 9 NO.1*, 9(1), 2025.
<https://doi.org/10.47002/metik.v9i1.1055>