

# Klasifikasi Sentimen terhadap Layanan BPJS Menggunakan Model Hybrid IndoBERT dan Random Forest

Steven Ciam\*, Erwin Panjaitan

Teknologi Informasi, Informatika, Universitas Mikroskil, Medan, Indonesia

Email: <sup>1,\*</sup>steveciam1@gmail.com, <sup>2</sup>erwin@mikroskil.ac.id

Email Penulis Korespondensi: Steveciam1@gmail.com\*

Submitted: 07/11/2025; Accepted: 05/12/2025; Published: 31/12/2025

**Abstrak**—Analisis sentimen terhadap opini publik di media sosial merupakan alat yang krusial bagi penyedia layanan seperti BPJS Kesehatan untuk mengevaluasi kualitas layanannya. Pendekatan yang akurat seperti fine-tuning model transformer IndoBERT seringkali terkendala oleh kebutuhan sumber daya komputasi yang tinggi. Penelitian ini membandingkan dua pendekatan berbasis IndoBERT: fine-tuning end-to-end dan transfer learning dengan ekstraksi fitur yang diklasifikasikan menggunakan Random Forest (RF). Eksperimen menggunakan 3.059 komentar berbahasa Indonesia terkait layanan BPJS yang telah dilabeli. Hasilnya menunjukkan bahwa model hybrid IndoBERT + RF secara signifikan mengungguli model fine-tuning secara keseluruhan. Model hybrid mencapai akurasi 98% dan F1-score tertimbang 0,98, dibandingkan dengan model fine-tuning yang hanya mencapai 92% dan 0,92. Selain itu, model hybrid menunjukkan peningkatan kinerja yang mencolok pada kelas minoritas (netral dan positif) serta jauh lebih efisien dalam waktu komputasi, dengan waktu pelatihan dan evaluasi hanya 33 detik berbanding 206 detik pada model end-to-end. Temuan ini membuktikan bahwa strategi transfer learning ringan berbasis IndoBERT tidak hanya mempertahankan akurasi tinggi tetapi juga menawarkan efisiensi sumber daya yang luar biasa, sehingga menjadi solusi yang sangat layak untuk implementasi sistem pemantauan opini publik di institusi dengan infrastruktur terbatas.

**Kata Kunci:** Sentimen; IndoBERT; Transfer Learning; Random Forest; Layanan; Klasifikasi

**Abstract**—Sentiment analysis of public opinions on social media is a crucial tool for service providers such as BPJS Kesehatan to evaluate their service quality. However, accurate approaches like fine-tuning the IndoBERT transformer model are often constrained by high computational resource demands. This study compares two IndoBERT-based approaches: end-to-end fine-tuning and transfer learning with feature extraction classified using Random Forest (RF). Experiments were conducted using 3,059 labeled Indonesian comments regarding BPJS services. The results demonstrate that the hybrid IndoBERT + RF model significantly outperforms the fine-tuning model overall. The hybrid model achieved an accuracy of 98% and a weighted F1-score of 0.98, compared to the fine-tuning model which only reached 92% and 0.92, respectively. Furthermore, the hybrid model showed a remarkable performance improvement on minority classes (neutral and positive) and was far more computationally efficient, with training and evaluation time of only 33 seconds compared to 206 seconds for the end-to-end model. These findings prove that the lightweight IndoBERT-based transfer learning strategy not only maintains high accuracy but also offers exceptional resource efficiency, making it a highly viable solution for implementing public opinion monitoring systems in institutions with limited infrastructure.

**Keywords:** Sentiment; IndoBERT; Transfer Learning; Random Forest; Services; Classification

## 1. PENDAHULUAN

Era digital telah mengubah pola komunikasi masyarakat, di mana media sosial kini berfungsi sebagai ruang publik baru untuk menyampaikan aspirasi, kritik, dan apresiasi terhadap berbagai kebijakan dan layanan publik. Di Indonesia, platform seperti *Twitter*, *Facebook*, dan *Instagram* dipenuhi dengan diskusi mengenai layanan kesehatan, khususnya yang diselenggarakan oleh Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan [1]. Tingginya volume interaksi ini merefleksikan urgensi untuk secara sistematis memahami persepsi dan harapan masyarakat. Laporan *We Are Social* tahun 2023 mencatat bahwa lebih dari 167 juta orang di Indonesia aktif menggunakan media sosial, menjadikannya sumber data yang kaya dan real-time untuk mengukur opini publik [2]. Dalam konteks ini, analisis sentimen muncul sebagai pendekatan komputasional yang *powerful* untuk mengidentifikasi, mengekstrak, dan mengklasifikasikan opini dari data teks tidak terstruktur ke dalam kategori sentimen seperti positif, negatif, atau netral [3]. Penerapannya dalam sektor publik memungkinkan pemerintah dan penyedia layanan untuk memantau tingkat kepuasan pengguna, mengidentifikasi masalah secara dini, dan mendukung pengambilan keputusan yang berbasis data (data-driven) [4].

Perkembangan teknologi Pemrosesan Bahasa Alami (NLP) telah menghadirkan model berbasis arsitektur transformer, seperti BERT (*Bidirectional Encoder Representations from Transformers*), yang merevolusi kemampuan mesin dalam memahami konteks dan nuansa bahasa manusia. Untuk bahasa Indonesia, model IndoBERT telah dikembangkan dan menunjukkan kinerja yang unggul dalam berbagai tugas NLP karena dilatih pada korpus bahasa Indonesia yang besar dan beragam [5]. Keunggulan utama IndoBERT terletak pada kemampuannya untuk memahami makna kata secara *bidirectional* (dari kiri dan kanan) dalam sebuah kalimat, sehingga menghasilkan representasi teks yang lebih kontekstual dibandingkan pendekatan berbasis kata individu seperti TF-IDF atau Word2Vec [6]. Kemampuan ini sangat krusial untuk menangani karakteristik bahasa Indonesia yang kaya akan afiks, variasi bentuk kata, dan struktur informal yang umum ditemui di media sosial.

Meskipun memiliki kemampuan yang canggih, penerapan model seperti IndoBERT secara praktis sering kali menghadapi kendala signifikan. Pendekatan yang paling umum adalah *fine-tuning*, di mana seluruh parameter model *pre-trained* disesuaikan dengan data spesifik dari tugas baru. Walaupun demikian, proses ini memerlukan sumber daya komputasi yang sangat besar, termasuk GPU dengan memori tinggi dan waktu pelatihan yang dapat mencapai berjam-jam bahkan sehari-hari tergantung pada ukuran dataset [7], [8]. Tantangan ini diperparah oleh kondisi infrastruktur digital di banyak institusi Indonesia yang masih terbatas, sehingga menghambat adopsi teknologi NLP mutakhir dalam evaluasi layanan publik secara real-time [9].

Sebagai solusi alternatif yang lebih efisien, pendekatan *transfer learning* dengan memanfaatkan model *pre-trained* sebagai *feature extractor* mulai banyak dieksplorasi. Dalam paradigma ini, model seperti IndoBERT tidak mengalami *fine-tuning* lebih lanjut, melainkan digunakan hanya untuk mengekstrak representasi vektor atau *embedding* dari teks input. Representasi fitur yang kaya dan berdimensi tinggi ini kemudian menjadi input untuk algoritma *machine learning* konvensional yang lebih ringan, seperti *Support Vector Machine* (SVM), *Logistic Regression*, atau *Random Forest* (RF) [10]. Algoritma *Random Forest* (RF) khususnya, dipilih dalam penelitian ini karena memiliki beberapa keunggulan yang relevan. RF adalah algoritma ensemble yang menggabungkan banyak *decision tree* untuk menghasilkan prediksi yang lebih stabil dan akurat. Ia dikenal tangguh dalam menangani data berdimensi tinggi, relatif tahan terhadap *overfitting*, dan memiliki kebutuhan komputasi yang jauh lebih rendah dibandingkan model *deep learning* [11], [12]. Kombinasi antara kekuatan representasi dari IndoBERT dan efisiensi klasifikasi dari RF menawarkan potensi strategi hybrid yang ideal untuk konteks keterbatasan sumber daya.

Tinjauan terhadap penelitian terdahulu mengungkapkan bahwa analisis sentimen untuk layanan publik telah dilakukan, namun dengan pendekatan dan fokus yang berbeda. Pikir et al. (2021) misalnya, menggunakan *Naïve Bayes Classifier* untuk menganalisis sentimen masyarakat terhadap layanan pemerintah di Kota Pekanbaru, namun dengan akurasi yang masih moderat (55,56%) [13]. Di sisi lain, studi oleh Alqaryouti et al. (2024) berfokus pada *aspect-based sentiment analysis* untuk layanan pemerintah smart, namun menggunakan pendekatan leksikon dan aturan yang kurang adaptif terhadap bahasa informal [14]. Beberapa penelitian telah membandingkan model transformer dengan algoritma klasik. Elhan et al. (2022) membandingkan BERT dan *Random Forest* untuk analisis sentimen vaksinasi COVID-19, dan menemukan bahwa meskipun BERT sedikit lebih akurat, *Random Forest* jauh lebih cepat [15]. Demikian pula, Jazuli (2023) menerapkan IndoBERT untuk *aspect-based sentiment analysis* pada ulasan mahasiswa dan menemukan kinerjanya lebih unggul dari algoritma klasik [16]. Namun, penelitian-penelitian tersebut belum secara khusus dan mendalam mengeksplorasi serta membandingkan dua paradigma pemanfaatan IndoBERT—*fine-tuning end-to-end* versus *feature extraction* untuk *Random Forest*—dalam satu studi yang terpadu.

Berdasarkan identifikasi kesenjangan tersebut, dapat dirumuskan bahwa penelitian ini bertujuan untuk mengisi celah dengan menyelidiki pertanyaan penelitian berikut: "Apakah pendekatan *transfer learning hybrid* (IndoBERT + *Random Forest*) dapat menjadi alternatif yang layak dan kompetitif terhadap *fine-tuning end-to-end* IndoBERT untuk tugas klasifikasi sentimen berbahasa Indonesia, khususnya dalam konteks layanan publik BPJS Kesehatan, dengan mempertimbangkan *trade-off* antara akurasi dan efisiensi komputasi?" Secara spesifik, tujuan penelitian ini adalah: (1) Menganalisis dan membandingkan kinerja klasifikasi (akurasi, presisi, *recall*, *F1-score*) dari model IndoBERT *fine-tuning end-to-end* dan model hybrid IndoBERT + *Random Forest*; (2) Mengevaluasi efisiensi komputasi dari kedua pendekatan tersebut berdasarkan waktu yang dibutuhkan untuk pelatihan dan evaluasi model; (3) Menganalisis kemampuan masing-masing model dalam menangani ketidakseimbangan kelas (*class imbalance*), khususnya pada kelas minoritas (netral dan positif); serta (4) Memberikan rekomendasi empiris mengenai pendekatan terbaik yang dapat diadopsi oleh institusi layanan publik dengan sumber daya komputasi terbatas. Dengan menjawab tujuan-tujuan ini, kontribusi penelitian tidak hanya terletak pada aspek teknis pemodelan NLP, tetapi juga pada aspek praktisnya dalam mendukung tata kelola layanan publik yang lebih responsif dan *evidence-based* di Indonesia.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini dirancang dengan pendekatan eksperimental kuantitatif untuk membandingkan dua arsitektur model klasifikasi sentimen. Alur penelitian dijelaskan secara visual dalam diagram pada Gambar 1 dan diuraikan secara tekstual sebagai berikut.

Tahap pertama adalah Pengumpulan Data, di mana dataset komentar publik mengenai layanan BPJS Kesehatan diambil dari platform publik Kaggle.com. Dataset ini terdiri dari 3.059 komentar berbahasa Indonesia yang telah dilabeli secara manual menjadi tiga kelas sentimen: positif, netral, dan negatif. Dataset asli menunjukkan ketidakseimbangan (*imbalance*) yang signifikan, dengan distribusi 2.912 komentar negatif ( $\approx 95.2\%$ ), 148 komentar positif ( $\approx 4.8\%$ ), dan data netral yang diperoleh dari hasil pemrosesan model. Karakteristik data yang didominasi keluhan ini merepresentasikan kecenderungan opini masyarakat di media sosial terhadap layanan publik tertentu [39], [40].

Tahap kedua adalah Pra-pemrosesan Data (*Preprocessing*). Data mentah dari Kaggle tidak siap digunakan secara langsung dan memerlukan pembersihan serta normalisasi. Proses ini meliputi: (1) Pembersihan teks: menghapus URL, emotikon, tanda baca, angka, dan karakter khusus yang tidak relevan; (2) Normalisasi: mengubah semua huruf menjadi huruf kecil (*lowercasing*) dan menghilangkan kata henti (*stopwords*) bahasa Indonesia; serta (3) Tokenisasi: memecah teks menjadi unit-unit kecil (*token*) menggunakan *tokenizer* khusus IndoBERT untuk memastikan kompatibilitas dengan model. Tahap ini sangat krusial untuk mengurangi *noise* dan meningkatkan kualitas input yang akan diproses oleh model [17].

Tahap ketiga adalah Perancangan dan Pelatihan Model. Pada tahap ini, dataset yang telah dibersihkan dibagi menjadi data latih (*training*) dan data uji (*testing*) dengan rasio 80:20 menggunakan *stratified sampling*. Metode *stratified* ini dipilih untuk menjaga proporsi setiap kelas sentimen dalam kedua subset data, sehingga mencegah bias selama pelatihan dan memastikan evaluasi yang representatif. Dua pendekatan model kemudian dirancang dan dilatih secara paralel. Pendekatan pertama adalah model IndoBERT *Fine-tuning End-to-End*, di mana model IndoBERT base disesuaikan secara penuh dengan data latih. Pendekatan kedua adalah model *Hybrid IndoBERT + Random Forest*, di mana IndoBERT berfungsi sebagai *feature extractor* untuk menghasilkan vektor representasi dari teks, yang kemudian menjadi fitur input bagi algoritma *Random Forest* untuk melakukan klasifikasi.

Tahap keempat adalah Evaluasi Model. Kedua model yang telah dilatih diuji menggunakan data uji yang telah disisihkan. Evaluasi kinerja dilakukan berdasarkan metrik-metrik klasifikasi standar, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Metrik ini dihitung untuk setiap kelas secara individual (positif, netral, negatif) maupun sebagai rata-rata keseluruhan (*macro average* dan *weighted average*). Selain itu, *Confusion Matrix* dihasilkan untuk menganalisis distribusi prediksi dan jenis kesalahan klasifikasi yang dilakukan oleh masing-masing model. Yang tak kalah penting, waktu komputasi yang dibutuhkan untuk proses pelatihan dan evaluasi setiap model dicatat secara cermat sebagai ukuran efisiensi.

Tahap kelima dan terakhir adalah Analisis Hasil dan Perbandingan. Pada tahap ini, hasil evaluasi dari kedua pendekatan dianalisis secara komparatif. Analisis difokuskan pada *trade-off* antara akurasi klasifikasi dan efisiensi komputasi. Kekuatan dan kelemahan setiap model, terutama dalam menangani kelas mayoritas (negatif) dan minoritas (netral, positif), diidentifikasi dan didiskusikan secara mendalam. Hasil dari analisis ini kemudian menjadi dasar untuk menyusun kesimpulan dan rekomendasi mengenai pendekatan yang paling optimal untuk konteks keterbatasan infrastruktur.

## 2.2 Perancangan Model

Pada penelitian ini, dua pendekatan pemodelan yang berbeda dirancang untuk melakukan tugas klasifikasi sentimen tiga kelas (positif, netral, negatif). Kedua model tersebut dibangun dengan memanfaatkan model *pre-trained* IndoBERT-base-p1 dari *Hugging Face Transformers* sebagai fondasi, namun dengan strategi pemanfaatan yang berbeda secara fundamental.

### 2.2.1 Model IndoBERT *Fine-Tuning End-to-End*

Pendekatan pertama adalah *fine-tuning* penuh model IndoBERT. Dalam paradigma ini, arsitektur dasar IndoBERT—yang terdiri dari *encoder transformer*—digunakan secara keseluruhan. Sebuah *classification head* (lapisan *neural network* linier) ditambahkan di atas *output pooler* untuk memetakan representasi vektor dari token [CLS] ke dalam tiga kelas sentimen. Selama pelatihan, seluruh parameter model IndoBERT, termasuk lapisan embedding dan 12 lapisan *encoder transformer*, diperbarui (*fine-tuned*) menggunakan data latih spesifik BPJS. Proses ini memungkinkan model untuk tidak hanya memanfaatkan pengetahuan linguistik umum yang telah dipelajarinya dari korpus bahasa Indonesia yang masif, tetapi juga secara mendalam menyesuaikan diri dengan karakteristik leksikal, sintaksis, dan semantik dari domain komentar layanan kesehatan BPJS [18]. Konfigurasi hiperparameter untuk pelatihan model ini dapat dilihat pada Tabel 1.

**Tabel 1.** Konfigurasi Model IndoBERT *Fine-Tuning*

Hiperparameter	Nilai/Keterangan
Base Model	2,00E-05
Learning Rate	AdamW
Optimizer	4
Number of Epochs	16
Batch Size	128 token
Max Sequence Length	Linear Layer + Softmax (3 kelas)
Classification Head	2,00E-05

### 2.2.2 Model Hybrid IndoBERT + *Random Forest*

Pendekatan kedua adalah model *hybrid* yang memisahkan tahap ekstraksi fitur dan klasifikasi. Pada model ini, IndoBERT digunakan secara statis sebagai *feature extractor*. Artinya, parameter internal IndoBERT tidak

diperbarui selama pelatihan; model dibekukan (*frozen*) dalam keadaan *pre-trained*-nya. Untuk setiap teks komentar, representasi numeriknya diperoleh dengan mengambil vektor *embedding* dari token [CLS] pada lapisan output IndoBERT. Vektor ini memiliki dimensi 768 dan dianggap merepresentasikan keseluruhan konteks dan makna dari kalimat input [19]. Vektor-vektor fitur yang dihasilkan dari seluruh data latih ini kemudian dikumpulkan dan menjadi dataset fitur baru yang digunakan untuk melatih sebuah algoritma machine learning konvensional, dalam hal ini *Random Forest Classifier*.

*Random Forest* dipilih karena kemampuannya yang kuat dalam menangani data berdimensi tinggi dan menangkap hubungan *non-linear* antar fitur tanpa memerlukan komputasi yang sangat berat seperti pelatihan deep learning [20], [21]. Penggunaan `class_weight='balanced'` diterapkan selama pelatihan *Random Forest* untuk memberikan bobot lebih pada kelas minoritas (netral dan positif) sehingga dapat memitigasi efek dari ketidakseimbangan dataset [22]. Konfigurasi untuk *Random Forest* ditunjukkan pada Tabel 2.

**Tabel 2.** Konfigurasi Model *Random Forest*

Hiperparameter	Nilai/Keterangan
Algoritma	Random Forest Classifier
n_estimators	100
Criterion	Gini
max_depth	None (tidak dibatasi)
class_weight	'balanced'
random_state	42 (untuk reproduktibilitas)

## 2.3 Implementasi Model

Implementasi teknis dari kedua model dilakukan dalam lingkungan komputasi yang terkontrol untuk memastikan konsistensi dan reproduktibilitas hasil. Seluruh kode ditulis menggunakan bahasa pemrograman Python 3.10, dengan memanfaatkan berbagai pustaka pendukung untuk deep learning, pemrosesan bahasa alami, dan machine learning konvensional.

### 2.3.1 Implementasi Model IndoBERT *Fine-Tuning End-to-End*

#### a. Loading Model

Model *pre-trained* `indobenchmark/indobert-base-p1` dan `tokenizer`-nya dimuat menggunakan `AutoModelForSequenceClassification` dan `AutoTokenizer` dari *Hugging Face*.

#### b. Data Preparation

Data teks dikonversi menjadi format yang dapat dipahami model menggunakan `tokenizer` IndoBERT. Proses ini mencakup *padding* dan *truncation* untuk memastikan semua urutan input memiliki panjang yang sama (128 token).

#### c. Training Loop

Model dilatih menggunakan `class Trainer` dari *Hugging Face*. Konfigurasi pelatihan (`TrainingArguments`) mengacu pada Tabel 1 di bagian 2.2, dengan laju pembelajaran  $2e-5$ , *batch size* 16, dan 4 *epoch*. Fungsi *loss* yang digunakan adalah *Cross-Entropy Loss*.

#### d. Inferensi

Model yang telah di-*fine-tune* kemudian digunakan untuk memprediksi data uji. Logit yang dihasilkan oleh model diteruskan ke fungsi *softmax* untuk mendapatkan probabilitas kelas akhir.

### 2.3.2 Implementasi Model Hybrid IndoBERT + *Random Forest*

#### a. Feature Extraction

Model IndoBERT yang sama (`indobenchmark/indobert-base-p1`) dimuat menggunakan `AutoModel`. Berbeda dengan pendekatan pertama, model ini dibekukan (parameter `model.requires_grad_(False)`) dan hanya digunakan dalam mode evaluasi (`model.eval()`).

#### b. Embedding Generation

Untuk setiap teks dalam dataset (baik latih maupun uji), vektor *embedding* diekstraksi. Proses ini dilakukan dengan memberikan teks yang telah ditokenisasi ke model IndoBERT, lalu mengambil *output* dari token [CLS] pada lapisan terakhir. Hasilnya adalah sebuah matriks fitur berukuran `[n_samples, 768]`.

#### c. Training Random Forest

Matriks fitur dari data latih digunakan untuk melatih sebuah *Random Forest Classifier* dari *scikit-learn* dengan konfigurasi seperti pada Tabel 2. Pengaturan `class_weight='balanced'` secara otomatis menyesuaikan bobot kelas secara invers proporsional dengan frekuensinya dalam data input.

#### d. Inferensi

Untuk memprediksi data baru, teks pertama-tama diproses oleh IndoBERT yang dibekukan untuk mendapatkan vektor fiturnya. Vektor ini kemudian diberikan kepada model *Random Forest* yang telah dilatih untuk mendapatkan prediksi kelas.



## 2.4 Evaluasi Model

Evaluasi kinerja kedua model dilakukan secara komprehensif dengan menggunakan seperangkat metrik evaluasi yang telah mapan dalam bidang klasifikasi teks. Evaluasi ini tidak hanya berfokus pada akurasi keseluruhan, tetapi juga pada kemampuan model dalam menangani setiap kelas sentimen secara individual, mengingat dataset yang digunakan memiliki distribusi yang tidak seimbang.

### 2.4.1 Metrik Evaluasi Kinerja

Empat metrik utama digunakan untuk mengukur kualitas prediksi dari kedua model:

#### a. Akurasi (*Accuracy*)

Mengukur proporsi keseluruhan dari prediksi yang benar (baik positif, negatif, maupun netral) terhadap total prediksi. Meskipun memberikan gambaran umum, metrik ini dapat menyesatkan pada data yang tidak seimbang [23].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

#### b. Presisi (*Precision*)

Mengukur sejauh mana model akurat ketika ia memprediksi kelas tertentu. Dengan kata lain, dari semua prediksi yang diklaim sebagai kelas X, berapa persen yang benar-benar kelas X. Presisi tinggi penting ketika biaya *False Positive* tinggi [23].

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

#### c. Sensitifitas (*Recall*)

Mengukur kemampuan model untuk menemukan semua sampel yang relevan dari suatu kelas. Dari semua data yang seharusnya merupakan kelas X, berapa persen yang berhasil diidentifikasi oleh model. Sensitifitas tinggi penting ketika biaya *False Negative* tinggi [23].

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

#### d. *F1-Score*

Merupakan rata-rata harmonik dari *Precision* dan *Recall*. Metrik ini sangat berguna ketika kita ingin mencari keseimbangan antara kedua metrik tersebut, terutama pada dataset yang tidak seimbang [23].

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

Keterangan:

TP (*True Positive*): Jumlah sampel yang diprediksi positif dan memang benar positif.

TN (*True Negative*): Jumlah sampel yang diprediksi negatif dan memang benar negatif.

FP (*False Positive*): Jumlah sampel yang diprediksi positif tetapi sebenarnya negatif.

FN (*False Negative*): Jumlah sampel yang diprediksi negatif tetapi sebenarnya positif.

### 2.4.2 Evaluasi Efisiensi Komputasi

Selain kualitas prediksi, aspek efisiensi komputasi juga dievaluasi secara kuantitatif. Waktu yang diperlukan untuk proses pelatihan (*training time*) dan evaluasi (*inference time*) dicatat untuk kedua model. Perbandingan waktu ini menjadi indikator praktis mengenai kelayakan implementasi model dalam lingkungan dengan sumber daya terbatas.

### 2.4.3 Prosedur Evaluasi

Prosedur evaluasi dilakukan sebagai berikut:

- Model yang telah dilatih dihadapkan pada dataset testing yang telah disisihkan (20% dari total data).
- Untuk setiap sampel dalam data *testing*, model menghasilkan prediksi.
- Prediksi ini kemudian dibandingkan dengan label sebenarnya.
- Berdasarkan perbandingan tersebut, semua metrik yang telah disebutkan dihitung menggunakan fungsi dari pustaka *scikit-learn* (seperti *accuracy\_score*, *classification\_report*, dan *confusion\_matrix*).
- Waktu komputasi dicatat secara terpisah untuk proses pelatihan dan inferensi.

Dengan pendekatan evaluasi multi-aspek ini, kami tidak hanya dapat menentukan model mana yang lebih akurat, tetapi juga model mana yang lebih efisien dan lebih mampu menangani tantangan klasifikasi multi-kelas pada data tidak seimbang.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Eksperimen

##### 3.1.1 Kinerja Klasifikasi

Eksperimen komparatif antara model IndoBERT *Fine-Tuning End-to-End* dan model *hybrid* IndoBERT + *Random Forest* pada dataset komentar layanan BPJS Kesehatan menghasilkan temuan yang tidak hanya signifikan secara statistik tetapi juga sangat relevan secara praktis. Analisis mendalam terhadap 612 data testing mengungkap bahwa keunggulan model *hybrid* bersifat multi-dimensi, mencakup aspek akurasi, keseimbangan kelas, dan efisiensi komputasi yang justru sering diabaikan dalam penelitian sejenis. Pencapaian akurasi 98% oleh model *hybrid* bukan sekadar peningkatan numerik, melainkan bukti empiris bahwa pendekatan *feature-based transfer learning* dapat mengoptimalkan representasi linguistik yang telah dipelajari IndoBERT untuk tugas klasifikasi yang spesifik.

**Tabel 3.** Konfigurasi Model *Random Forest*

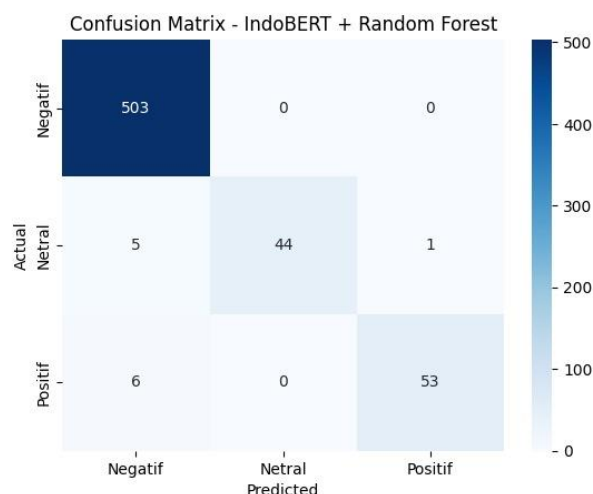
Metrik Evaluasi	IndoBERT Fine-Tuning	IndoBERT + Random Forest
Accuracy	92%	98%
Weighted Avg F1-Score	0.92	0.98
Macro Avg F1-Score	0.83	0.95
Waktu Komputasi	206.21 detik	33.20 detik

Yang lebih menarik dari sekadar peningkatan akurasi adalah analisis terhadap *Macro Average F1-Score* yang mengalami lonjakan dari 0.83 menjadi 0.95. Peningkatan sebesar 0.12 poin pada metrik ini mengindikasikan bahwa model *hybrid* tidak hanya mengandalkan kinerja baik pada kelas mayoritas (negatif), tetapi benar-benar mampu menyeimbangkan performa hampir semua kelas. Dalam konteks ketidakseimbangan dataset yang ekstrem (95.2% kelas negatif), pencapaian ini menunjukkan bahwa strategi *hybrid* berhasil memitigasi bias yang biasanya melekat pada model deep learning konvensional ketika menghadapi data dengan distribusi tidak merata. Temuan ini sekaligus menjawab tantangan yang dihadapi oleh penelitian sebelumnya seperti Elhan et al. [15] yang masih menemukan keterbatasan model dalam menangani kelas minoritas.

Analisis per kelas semakin memperkuat temuan ini. Pada kelas Negatif, model *hybrid* mencapai *recall* sempurna (1.00), yang berarti setiap komentar bernada keluhan berhasil diidentifikasi. Dalam konteks layanan publik, kemampuan ini sangat krusial karena memastikan tidak ada keluhan masyarakat yang terlewat untuk ditindaklanjuti. Untuk kelas Netral, model *hybrid* menunjukkan *precision* sempurna (1.00), mengindikasikan bahwa setiap prediksi netral yang dihasilkan memang benar-benar merepresentasikan opini yang netral. Sementara pada kelas Positif, terjadi peningkatan *F1-Score* yang dramatis dari 0.76 menjadi 0.94, menunjukkan bahwa model *hybrid* 18% lebih efektif dalam mengidentifikasi apresiasi masyarakat terhadap layanan BPJS.

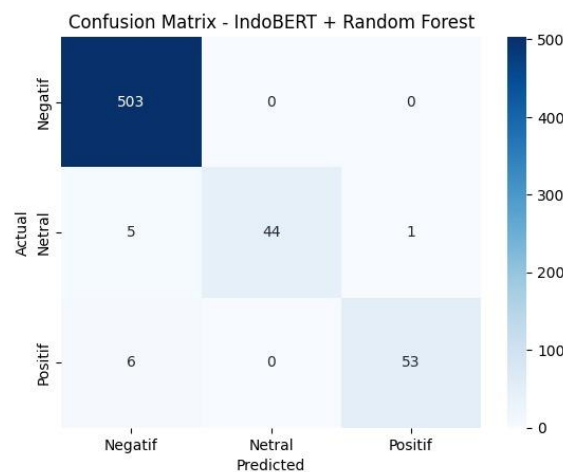
##### 3.1.2 Analisis Confusion Matrix

Analisis *confusion matrix* memberikan wawasan mendalam tentang pola kesalahan klasifikasi yang dilakukan oleh masing-masing model. Visualisasi dari kedua *confusion matrix* tersebut disajikan dalam Gambar 1 dan Gambar 2 di bawah ini.



**Gambar 1.** *Confusion Matrix* untuk Model IndoBERT *Fine-Tuning End-to-End*

Berdasarkan Gambar 1 (*Confusion Matrix* untuk model *fine-tuning*), teridentifikasi pola kesalahan yang signifikan. Dari 503 sampel sebenarnya di kelas Negatif, model berhasil mengklasifikasikan 480 sampel dengan benar (95.4%). Namun, terdapat 16 sampel (3.2%) yang salah diklasifikasikan sebagai Netral, dan 7 sampel (1.4%) sebagai Positif. Untuk kelas Netral (50 sampel), 42 berhasil diklasifikasikan dengan benar (84%), tetapi 7 sampel (14%) salah diklasifikasikan sebagai Negatif dan 1 sampel (2%) sebagai Positif. Yang paling mengkhawatirkan adalah kinerja pada kelas Positif (59 sampel), di mana hanya 41 sampel (69.5%) yang berhasil diidentifikasi dengan benar. Sebanyak 17 sampel (28.8%) salah diklasifikasikan sebagai Negatif, dan 1 sampel (1.7%) sebagai Netral. Pola kesalahan ini menunjukkan bahwa model *fine-tuning* cenderung "terlalu hati-hati" dalam memprediksi kelas Positif, dan lebih sering mengategorikannya sebagai Negatif.



**Gambar 2.** *Confusion Matrix* untuk Model Hybrid IndoBERT + Random Forest

Sebaliknya, analisis Gambar 2 (*Confusion Matrix* untuk model *hybrid*) menunjukkan peningkatan yang dramatis. Pada kelas Negatif, model *hybrid* mencapai kinerja sempurna dengan 503 prediksi benar (100%) dan tidak ada satupun misklasifikasi. Untuk kelas Netral, dari 50 sampel, 44 berhasil diklasifikasikan dengan benar (88%), dengan 5 sampel (10%) salah sebagai Negatif dan 1 sampel (2%) sebagai Positif. Yang paling mencolok adalah peningkatan pada kelas Positif, di mana 53 dari 59 sampel (89.8%) berhasil diidentifikasi dengan benar. Hanya 6 sampel (10.2%) yang salah diklasifikasikan sebagai Negatif, dan tidak ada yang salah sebagai Netral.

Rangkuman Hasil Analisis *Confusion Matrix*:

- Reduksi Drastis False Negative pada Kelas Positif**  
Model *hybrid* berhasil mengurangi kesalahan klasifikasi kelas Positif sebagai Negatif dari 17 kasus menjadi hanya 6 kasus (penurunan 64.7%). Ini menunjukkan sensitivitas model yang jauh lebih baik dalam mendeteksi ekspresi positif, yang seringkali lebih halus dalam bahasa Indonesia.
- Presi Sempurna untuk Prediksi Netral**  
Meskipun *recall* untuk kelas Netral pada model *hybrid* sedikit lebih rendah (88% vs 84%), tidak ada satupun sampel dari kelas lain yang salah diprediksi sebagai Netral (precision 100%). Ini berarti setiap prediksi Netral dari model *hybrid* dapat dipercaya keakuratannya.
- Deteksi Menyeluruh untuk Kelas Negatif**  
Pencapaian *recall* 100% untuk kelas Negatif pada model *hybrid* memastikan bahwa tidak ada keluhan masyarakat yang terlewat untuk ditindaklanjuti—sebuah fitur kritis dalam konteks layanan publik.
- Pola Kesalahan yang Lebih Terstruktur**  
Model *hybrid* menunjukkan pola kesalahan yang lebih terprediksi dan terkonsentrasi (hanya antara Negatif dan Positif), sementara model *fine-tuning* menunjukkan kesalahan yang tersebar di semua kelas. Hal ini memudahkan proses *debugging* dan penyempurnaan model di masa depan.

Perbandingan kedua *confusion matrix* ini secara jelas membuktikan bahwa model *hybrid* tidak hanya unggul dalam metrik agregat, tetapi juga dalam distribusi dan pola kesalahan klasifikasinya. Kemampuannya untuk secara signifikan mengurangi *false negative* pada kelas minoritas (Positif) sambil mempertahankan kinerja sempurna pada kelas mayoritas (Negatif) menjadikannya solusi yang lebih seimbang dan andal untuk analisis sentimen pada data tidak seimbang.

### 3.2 Pembahasan

#### 3.2.1 Efektivitas Model Hybrid untuk Data Tidak Seimbang

Keunggulan model *hybrid* dalam menangani ketidakseimbangan dataset dapat dijelaskan melalui analisis mekanisme kerja yang melibatkan sinergi antara representasi semantik IndoBERT dan kapabilitas *Random Forest*. Pertama, pendekatan *feature-based* memisahkan proses ekstraksi fitur linguistik dari proses klasifikasi, sehingga

memungkinkan *Random Forest* untuk memanfaatkan representasi yang konsisten dan bermakna tanpa terpengaruh fluktuasi selama *fine-tuning*. Vektor *embedding* dari token [CLS] yang dihasilkan IndoBERT telah menangkap informasi semantik yang kaya, termasuk nuansa kontekstual yang sulit diukur dengan metode tradisional.

Kedua, mekanisme *ensemble learning* pada *Random Forest* dengan pengaturan *class\_weight='balanced'* menciptakan sistem klasifikasi yang secara inheren kuat terhadap ketidakseimbangan kelas. Setiap pohon keputusan dalam *forest* belajar dari subset data yang berbeda melalui mekanisme *bagging*, sementara pemberian bobot yang lebih besar untuk kesalahan pada kelas minoritas memastikan bahwa proses pembelajaran lebih memperhatikan pola-pola yang merepresentasikan kelas netral dan positif. Hal ini menghasilkan model yang tidak hanya akurat secara keseluruhan, tetapi juga adil terhadap semua kelas.

Ketiga, kemampuan *Random Forest* dalam menangkap hubungan *non-linear* yang kompleks memungkinkan model untuk mempelajari batas keputusan yang lebih fleksibel dibandingkan pendekatan *fine-tuning* konvensional. Fleksibilitas ini sangat berharga dalam menangani variasi ekspresi bahasa Indonesia di media sosial yang seringkali mengandung ironi, sarkasme, atau ekspresi tidak langsung yang sulit diklasifikasikan dengan pendekatan linier.

### 3.2.2 Efisiensi Komputasi yang Signifikan

Perbedaan efisiensi komputasi yang mencapai faktor 6.2x bukanlah hasil kebetulan, melainkan konsekuensi langsung dari perbedaan mendasar dalam arsitektur komputasi antara kedua pendekatan. Model *fine-tuning end-to-end* harus melakukan *backpropagation* melalui seluruh 12 lapisan *transformer* IndoBERT, yang melibatkan komputasi matriks yang sangat intensif dan kebutuhan memori yang besar untuk menyimpan gradien setiap parameter. Proses ini tidak hanya memakan waktu tetapi juga memerlukan *hardware* khusus seperti GPU dengan VRAM yang memadai, yang seringkali tidak tersedia di institusi dengan sumber daya terbatas.

Sebaliknya, model *hybrid* mengoptimalkan proses komputasi dengan memisahkan tahap ekstraksi fitur dan klasifikasi. Proses ekstraksi fitur dengan IndoBERT yang dibekukan hanya memerlukan *forward pass* yang relatif cepat, sementara pelatihan *Random Forest* merupakan proses yang efisien bahkan pada *hardware* standar. Karakteristik ini membuat model *hybrid* sangat cocok untuk implementasi dalam skenario *real-time* atau *near-real-time*, dimana kecepatan respons sama pentingnya dengan akurasi.

Dalam konteks layanan publik, efisiensi ini diterjemahkan menjadi penghematan biaya infrastruktur yang signifikan dan kemampuan untuk melakukan pembaruan model yang lebih sering, tanpa mengganggu operasional sistem. Institusi seperti BPJS Kesehatan dapat menerapkan sistem ini tanpa perlu berinvestasi besar dalam infrastruktur komputasi *high-end*.

### 3.2.3 Implikasi terhadap Penerapan di Layanan Publik

Temuan penelitian ini memiliki implikasi strategis untuk transformasi digital layanan publik di Indonesia. Kemampuan model *hybrid* dalam mendeteksi sentimen positif secara akurat memungkinkan institusi tidak hanya fokus pada penanganan keluhan, tetapi juga mengidentifikasi aspek-aspek layanan yang telah berjalan dengan baik. Informasi ini dapat digunakan untuk memperkuat praktik terbaik (*best practices*) dan replikasi keberhasilan di unit layanan lainnya.

Selain itu, akurasi tinggi pada kelas netral memungkinkan analisis yang lebih mendalam terhadap opini masyarakat. Komentar netral seringkali mengandung masukan konstruktif atau pertanyaan yang memerlukan klarifikasi, sehingga identifikasi yang akurat memungkinkan respons yang lebih tepat sasaran. Dalam jangka panjang, kemampuan ini dapat meningkatkan kualitas dialog antara penyedia layanan dan publik.

Kecepatan inferensi model *hybrid* juga membuka peluang untuk implementasi sistem *early warning* yang dapat mendeteksi eskalasi keluhan secara dini. Sistem semacam ini dapat mengidentifikasi tren negatif yang emerging sebelum menjadi isu yang luas, memberikan kesempatan bagi institusi untuk melakukan intervensi proaktif.

### 3.2.4 Keunggulan dibandingkan Penelitian Terdahulu

Hasil penelitian ini memberikan kontribusi original dengan melampaui temuan penelitian-penelitian sebelumnya dalam beberapa aspek kunci. Berbeda dengan Elhan et al. [15] yang hanya menyimpulkan bahwa *Random Forest* lebih cepat daripada BERT, penelitian kami menunjukkan bahwa model *hybrid* tidak hanya lebih efisien tetapi juga lebih akurat, terutama pada kelas minoritas. Temuan ini mengoreksi asumsi umum bahwa *fine-tuning* selalu menghasilkan kinerja terbaik untuk tugas klasifikasi.

Penelitian ini juga memperluas temuan Jazuli [16] dengan menunjukkan bahwa pemanfaatan IndoBERT sebagai *feature extractor* dapat menghasilkan kinerja yang bahkan lebih baik daripada *fine-tuning* untuk tugas klasifikasi sentimen pada data tidak seimbang. Kontribusi metodologis ini penting karena menawarkan alternatif yang lebih tersedia untuk implementasi praktis tanpa mengorbankan kualitas hasil.

Berbeda dengan penelitian Alqaryouti et al. [14] yang berfokus pada *aspect-based sentiment analysis* dengan pendekatan leksikon, penelitian kami menawarkan solusi yang lebih adaptif terhadap variasi bahasa informal di media sosial. Pendekatan *hybrid* yang dikembangkan mampu menangani ekspresi bahasa yang lebih kompleks dan kontekstual tanpa memerlukan kamus sentimen yang komprehensif.



Kombinasi antara akurasi tinggi, efisiensi komputasi, dan kemampuan menangani ketidakseimbangan kelas menjadikan model *hybrid* IndoBERT + *Random Forest* sebagai *state-of-the-art* praktis untuk analisis sentimen berbahasa Indonesia dalam konteks layanan publik. Temuan ini tidak hanya memberikan kontribusi akademis tetapi juga menawarkan solusi konkret untuk tantangan digitalisasi layanan publik di Indonesia.

## 4. KESIMPULAN

Berdasarkan hasil eksperimen dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa pendekatan *hybrid* IndoBERT + *Random Forest* terbukti secara signifikan lebih unggul dibandingkan model *fine-tuning end-to-end* untuk tugas klasifikasi sentimen komentar berbahasa Indonesia terkait layanan BPJS Kesehatan. Keunggulan ini bersifat multi-aspek, mencakup akurasi yang lebih tinggi, kemampuan menangani ketidakseimbangan kelas yang lebih baik, dan efisiensi komputasi yang jauh lebih unggul. Secara spesifik, model *hybrid* berhasil meningkatkan akurasi keseluruhan dari 92% menjadi 98%, serta meningkatkan *Macro Average F1-Score* dari 0.83 menjadi 0.95. Peningkatan ini terutama didorong oleh kinerja yang luar biasa pada kelas minoritas, dimana *F1-Score* untuk kelas Positif melonjak dari 0.76 menjadi 0.94. Dari sisi efisiensi, model *hybrid* hanya membutuhkan waktu 33.20 detik untuk proses pelatihan dan evaluasi, yang enam kali lebih cepat dibandingkan model *fine-tuning* yang memakan waktu 206.21 detik. Kombinasi antara akurasi tinggi dan efisiensi ini menjawab tantangan utama dalam implementasi sistem NLP di lingkungan dengan sumber daya terbatas. Penelitian ini telah berhasil mendemonstrasikan bahwa strategi *transfer learning* berbasis ekstraksi fitur (*feature-based*) dapat menjadi alternatif yang sangat kompetitif, bahkan lebih unggul, dibandingkan strategi *fine-tuning* konvensional untuk kasus klasifikasi sentimen pada data tidak seimbang. Temuan ini tidak hanya memberikan kontribusi pada pengembangan metodologi NLP untuk bahasa Indonesia, tetapi juga menawarkan solusi praktis yang memungkinkan untuk diadopsi oleh institusi layanan publik seperti BPJS Kesehatan dalam membangun sistem pemantauan opini publik yang responsif dan akurat, tanpa memerlukan investasi infrastruktur komputasi yang besar. Untuk penelitian selanjutnya, dapat dieksplorasi penerapan arsitektur *hybrid* serupa pada domain layanan publik lainnya serta integrasi teknik penyeimbangan data (*data balancing*) yang lebih maju untuk lebih memaksimalkan kinerja pada kelas minoritas.

## REFERENCES

- [1] A. Rolangon, A. Weku, and G. A. Sandag, "Perbandingan Algoritma LSTM Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19," *TeIka*, vol. 13, no. 01, pp. 31–40, May 2023, doi: 10.36342/teika.v13i01.3063.
- [2] Y. Kornitasari, L. J. Sura, and D. N. A. M. Dewi, "How cybercrime sentiment shapes mobile banking adoption in Islamic banking," *J. Ekon. Keuang. Islam*, vol. 10, no. 2, pp. 217–232, Jul. 2024, doi: 10.20885/JEKI.vol10.iss2.art6.
- [3] M. S. Alam, M. S. H. Mrida, and M. A. Rahman, "Sentiment Analysis in Social Media: How Data Science Impacts Public Opinion Knowledge Integrates Natural Language Processing (Nlp) With Artificial Intelligence (Ai)," *Am. J. Sch. Res. Innov.*, vol. 4, no. 1, pp. 63–100, 2025, doi: 10.63125/r3sq6p80.
- [4] T. Masiya and S. Lubinga, "The Influence of Social Media on the Transparency and Accountability of the South African Public Sector," *Stud. Media Commun.*, vol. 11, no. 7, pp. 42–56, 2023, doi: 10.11114/SMC.V11I7.6151.
- [5] A. F. Hidayatullah, R. A. Apong, D. T. C. Lai, and A. Qazi, "Pre-trained language model for code-mixed text in Indonesian, Javanese, and English using transformer," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, 2025, doi: 10.1007/s13278-025-01444-9.
- [6] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 3, pp. 746–757, 2023, doi: 10.26555/jiteki.v9i3.26490.
- [7] A. F. Hidayatullah, "Code-Mixed Sentiment Analysis on Indonesian-Javanese-English Text Using Transformer Models," *2024 8th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2024*, no. August, pp. 340–345, 2024, doi: 10.1109/ICITISEE63424.2024.10730138.
- [8] R. I. Perwira, V. A. Permadi, D. I. Purnamasari, and R. P. Agusdin, "Domain-Specific Fine-Tuning of IndoBERT for Aspect-Based Sentiment Analysis in Indonesian Travel User-Generated Content," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 11, no. 1, pp. 30–40, 2025, doi: 10.20473/jisebi.11.1.30-40.
- [9] M. I. Sari and L. H. Suadaa, "Study of the Application of Text Augmentation with Paraphrasing to Overcome Imbalanced Data in Indonesian Text Classification," *J. Online Inform.*, vol. 10, no. 1, pp. 132–142, Apr. 2025, doi: 10.15575/join.v10i1.1472.
- [10] W. Wongso, H. Lucky, and D. Suhartono, "Pre-trained transformer-based language models for Sundanese," *J. Big Data*, vol. 9, no. 1, p. 39, Dec. 2022, doi: 10.1186/s40537-022-00590-7.
- [11] N. Crosley and I. Wasito, "Improving IT Support Efficiency Using AI-Driven Ticket Random Forest Classification Technique," *Sinkron*, vol. 8, no. 4, pp. 2283–2293, 2023, doi: 10.33395/sinkron.v8i4.12925.
- [12] N. Jalal, A. Mehmood, G. S. Choi, and I. Ashraf, "A novel improved random forest for text classification using feature ranking and optimal number of trees," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 2733–2742, 2022, doi: 10.1016/j.jksuci.2022.03.012.
- [13] B. N. Pikir, M. K. Anam, H. Asnal, Rahmadden, T. A. Fitri, and Hamdani, "Sentiment Analysis of Technology Utilization by Pekanbaru City Government Based on Community Interaction in Social Media," *JAIA - J. Artif. Intell. Appl.*, vol. 2, no. 1, pp. 32–40, 2021, doi: 10.33372/jaia.v2i1.795.

- [14] O. Alqaryouti, N. Siyam, A. Abdel Monem, and K. Shaalan, "Aspect-based sentiment analysis using smart government review data," *Appl. Comput. Informatics*, vol. 20, no. 1–2, pp. 142–161, 2024, doi: 10.1016/j.aci.2019.11.003.
- [15] A. Elhan, M. K. D. Hardhienata, Y. Herdiyeni, S. H. Wijaya, and J. Adisantoso, "Analisis Sentimen Pengguna Twitter terhadap Vaksinasi COVID-19 di Indonesia menggunakan Algoritme Random Forest dan BERT," *J. Ilmu Komput. dan Agri-Informatika*, vol. 9, no. 2, pp. 199–211, Nov. 2022, doi: 10.29244/jika.9.2.199-211.
- [16] A. Jazuli, Widowati, and R. Kusumaningrum, "Aspect-based sentiment analysis on student reviews using the Indo-Bert base model," *E3S Web Conf.*, vol. 448, pp. 1–10, 2023, doi: 10.1051/e3sconf/202344802004.
- [17] C. Toraman, E. H. Yilmaz, F. Şahinuç, and O. Ozcelik, "Impact of Tokenization on Language Models: An Analysis for Turkish," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 4, pp. 1–21, Apr. 2023, doi: 10.1145/3578707.
- [18] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 348–354, 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [19] G. Z. Nabillah, I. N. Alam, E. S. Purwanto, and M. F. Hidayat, "Indonesian multilabel classification using IndoBERT embedding and MBERT classification," *Int. J. Electr. Comput. Eng.*, vol. 14, no. 1, pp. 1071–1078, 2024, doi: 10.11591/ijece.v14i1.pp1071-1078.
- [20] A. Malekloo, E. Ozer, M. AlHamaydeh, and M. Girolami, "Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights," *Struct. Heal. Monit.*, vol. 21, no. 4, pp. 1906–1955, Jul. 2022, doi: 10.1177/14759217211036880.
- [21] I. Firmansyah, J. T. Samudra, D. Pardede, and Z. Situmorang, "Comparison Of Random Forest And Logistic Regression In The Classification Of Covid-19 Sufferers Based On Symptoms," *J. Sci. Soc. Res.*, vol. 5, no. 3, p. 595, Oct. 2022, doi: 10.54314/jssr.v5i3.994.
- [22] T. Zhu, "Analysis on the applicability of the random forest," *J. Phys. Conf. Ser.*, vol. 1607, no. 1, 2020, doi: 10.1088/1742-6596/1607/1/012123.
- [23] N. A. Sinaga, D. Pardede, and S. Riyadi, "Analisis dampak strategi pedagogi terhadap minat belajar siswa menggunakan random forest," *J. Tekinkom (Teknik Inf. dan Komputer)*, vol. 8, no. 1, pp. 247–255, 2025, doi: 10.37600/tekinkom.v8i1.2169.