

Perbandingan Kinerja Model Transformer Monolingual Dalam Analisis Sentimen Terhadap Website Coretax Pada Platform X

Angga Airul Rohman^{1, a)} Ike Verawati^{2, b)}

1)Program Studi Informatika,
Fakultas Ilmu Komputer,
Universitas Amikom Yogyakarta,
Jl. Padjajaran, Ring Road Utara, Yogyakarta 55281, Indonesia

Author Emails

^{a)} Corresponding author: anggaairulr@students.amikom.ac.id
^{b)} ikeverawati@amikom.ac.id

Abstract. The implementation of the Coretax system as a modernization of tax administration in Indonesia has generated diverse and substantial public responses on social media, particularly on the X platform. Monitoring public sentiment toward this system is crucial for evaluating its acceptance and effectiveness from the government's perspective. This study aims to compare the performance of three monolingual Transformer-based models—IndoBERT, IndoBERTweet, and IndoRoBERTa—in conducting sentiment analysis on public discourse related to the Coretax system. Data were collected by crawling the X platform using the Tweet Harvest library with Coretax-related keywords, resulting in 8,154 tweets. The dataset subsequently underwent text preprocessing and data augmentation to address class imbalance. The models were fine-tuned and evaluated using precision, recall, F1-score, and accuracy metrics. The experimental results indicate that IndoBERT outperforms the other models, achieving the highest accuracy of 0.8889. Meanwhile, IndoBERTweet and IndoRoBERTa achieved accuracies of 0.8850 and 0.8732, respectively. These findings suggest that the IndoBERT-based model is more effective in capturing taxation-related linguistic contexts, and thus is recommended as an automated tool for monitoring public perception for relevant governmental institutions.

Keywords:

Sentiment Analysis, Coretax, IndoBERT, IndoBERTweet, IndoRoBERTa

Abstraksi. Implementasi sistem Coretax sebagai pembaruan administrasi perpajakan di Indonesia memicu beragam respons publik yang signifikan di media sosial, khususnya platform X. Pemantauan sentimen publik terhadap sistem ini memiliki urgensi tinggi sebagai bahan evaluasi penerimaan dan efektivitas sistem bagi pemerintah. Penelitian ini bertujuan untuk membandingkan kinerja tiga model *Transformer* monolingual, yaitu IndoBERT, IndoBERTweet, dan IndoRoBERTa, dalam melakukan analisis sentimen terhadap isu tersebut. Data dikumpulkan melalui mekanisme *crawling* pada platform X menggunakan *library Tweet Harvest* dengan kata kunci terkait Coretax sebanyak 8154 data, yang kemudian melalui tahapan praproses teks dan augmentasi data untuk menangani ketidakseimbangan kelas. Model kemudian di-*fine-tune* dan dievaluasi menggunakan metrik *precision*, *recall*, *f1-score*, dan akurasi. Hasil eksperimen menunjukkan bahwa IndoBERT memberikan kinerja terbaik dibandingkan model lainnya, dengan pencapaian akurasi 0.8889. Sementara itu, IndoBERTweet dan IndoRoBERTa masing-masing mencatatkan akurasi sebesar 0,8850 dan 0,8732. Temuan ini mengindikasikan bahwa model berbasis IndoBERT lebih efektif dalam menangkap konteks linguistik terkait perpajakan, sehingga direkomendasikan sebagai alat pemantauan persepsi masyarakat otomatis bagi instansi terkait.

Kata Kunci:

Analisis Sentimen, Coretax, IndoBERT, IndoBERTweet, IndoRoBERTa

PENDAHULUAN

Perkembangan media sosial telah mengubah cara masyarakat menyampaikan opini terhadap berbagai isu, termasuk layanan publik dan kebijakan pemerintah. Platform seperti X (sebelumnya Twitter) memungkinkan pengguna untuk mengekspresikan pendapat secara cepat, terbuka, dan masif [1]. Di Indonesia, X merupakan salah satu platform media sosial dengan jumlah pengguna yang tinggi [2], sehingga opini yang disampaikan di platform tersebut berpotensi menjadi indikator penting dalam mengevaluasi respons publik terhadap kebijakan dan layanan yang disediakan oleh pemerintah [3].

Salah satu layanan digital pemerintah yang mendapat perhatian publik adalah website Coretax yang dikembangkan oleh Direktorat Jenderal Pajak sebagai bagian dari reformasi administrasi perpajakan. Sejak implementasinya, Coretax memunculkan berbagai tanggapan dari masyarakat di platform X, khususnya terkait kendala teknis, performa sistem, dan pengalaman pengguna [4]. Banyaknya opini yang tersebar dalam bentuk cuitan menjadikan data dari X sebagai sumber informasi yang relevan untuk menggambarkan persepsi publik secara komprehensif. Namun, volume data yang besar dan bersifat tidak terstruktur menimbulkan tantangan dalam proses analisis secara manual.

Pendekatan Natural Language Processing (NLP) menjadi solusi yang efektif untuk menganalisis opini publik dalam skala besar [5]. Salah satu metode NLP yang banyak digunakan adalah Bidirectional Encoder Representations from Transformers (BERT), yang mampu memahami konteks bahasa secara dua arah dan menghasilkan representasi teks yang lebih akurat [6]. Untuk bahasa Indonesia, telah dikembangkan beberapa model transformer monolingual berbasis BERT, seperti IndoBERT, IndoBERTweet, dan IndoRoBERTa, yang dirancang dengan karakteristik data dan bahasa Indonesia. Meskipun demikian, penelitian terdahulu terkait analisis sentimen terhadap website Coretax masih didominasi oleh metode konvensional, seperti Naive Bayes, yang memiliki keterbatasan dalam menangkap konteks semantik secara mendalam [4].

Selain itu, penelitian yang secara khusus membandingkan kinerja model transformer monolingual bahasa Indonesia, yaitu IndoBERT, IndoBERTweet, dan IndoRoBERTa, dalam konteks analisis sentimen terhadap layanan digital pemerintah masih sangat terbatas. Oleh karena itu, diperlukan kajian yang lebih mendalam untuk mengevaluasi performa ketiga model tersebut pada data opini publik yang bersumber dari platform X.

Berdasarkan permasalahan tersebut, pertanyaan penelitian dalam studi ini adalah: *bagaimana perbandingan kinerja model IndoBERT, IndoBERTweet, dan IndoRoBERTa dalam melakukan analisis sentimen terhadap website Coretax di platform X?* Adapun tujuan penelitian ini adalah untuk membandingkan kinerja ketiga model transformer monolingual bahasa Indonesia berdasarkan metrik evaluasi yang digunakan dalam analisis sentimen.

Penelitian ini berfokus pada evaluasi dan perbandingan kinerja model IndoBERT, IndoBERTweet, dan IndoRoBERTa yang telah dilakukan fine-tuning menggunakan dataset opini publik terhadap website Coretax di platform X. Kontribusi utama dari penelitian ini adalah memberikan gambaran empiris mengenai model transformer monolingual yang paling optimal untuk analisis sentimen layanan digital pemerintah di Indonesia, serta melengkapi keterbatasan penelitian sebelumnya yang masih menggunakan pendekatan konvensional.

TINJAUAN PUSTAKA

Analisis sentimen berbasis media sosial telah banyak digunakan untuk memahami opini publik terhadap berbagai isu sosial dan kebijakan pemerintah. Platform X (Twitter) menjadi salah satu sumber data utama karena sifatnya yang terbuka, real-time, dan mencerminkan respons masyarakat secara langsung. Beberapa penelitian terdahulu telah memanfaatkan data dari platform ini dengan pendekatan yang beragam.

Sitanggang melakukan analisis sentimen terhadap program makan siang gratis menggunakan algoritma Naïve Bayes dengan data cuitan dari platform X [1]. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat ke dalam kategori positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa metode Naïve Bayes mampu memberikan performa yang cukup baik dalam mengolah data teks berbahasa Indonesia, meskipun masih memiliki keterbatasan dalam memahami konteks kalimat yang kompleks. Hal ini disebabkan oleh karakteristik Naïve Bayes yang mengasumsikan independensi antar fitur kata.

Penelitian serupa dilakukan oleh Hutagalung yang menganalisis sentimen masyarakat terhadap kebijakan kenaikan harga bahan bakar minyak (BBM) menggunakan algoritma Naïve Bayes [2]. Fokus penelitian ini adalah mengukur kecenderungan sentimen publik terhadap kebijakan pemerintah melalui opini di X. Hasil penelitian

menunjukkan dominasi sentimen negatif, yang mencerminkan penolakan publik terhadap kebijakan tersebut. Namun, model yang digunakan masih mengalami kesulitan dalam membedakan sentimen yang bersifat sarkastik atau ambigu, sehingga akurasi analisis sangat bergantung pada kualitas preprocessing data.

Az-haari juga menerapkan metode Naïve Bayes untuk menganalisis sentimen boikot produk pro-Israel di Indonesia berdasarkan data cuitan di X [3]. Penelitian ini berhasil mengungkap pola sentimen publik yang cenderung negatif terhadap produk tertentu. Meskipun demikian, pendekatan berbasis probabilistik sederhana yang digunakan belum mampu menangkap hubungan semantik antar kata secara mendalam, terutama pada opini yang memiliki struktur kalimat kompleks.

Seiring dengan perkembangan NLP, pendekatan berbasis transformer mulai banyak digunakan karena kemampuannya dalam memahami konteks kalimat secara dua arah. Geni menerapkan model IndoBERT untuk analisis sentimen cuitan menjelang Pemilu 2024 [7]. Penelitian ini menggunakan dataset bertopik politik dan menunjukkan bahwa IndoBERT mampu mencapai akurasi sebesar 83,5% dan F1-score 88,49%, melampaui metode TextBlob, Naïve Bayes, dan SVM. Keunggulan ini disebabkan oleh kemampuan IndoBERT dalam mempelajari representasi kontekstual bahasa Indonesia, sehingga lebih efektif dalam menangkap makna opini publik yang bersifat kompleks.

Singgalen juga menggunakan IndoBERT untuk menganalisis sentimen ulasan hotel di Indonesia [8]. Hasil penelitian menunjukkan akurasi keseluruhan sebesar 92,5% dengan F1-score positif mencapai 96,1%. Namun, penelitian ini menemukan adanya bias pada kelas netral dan negatif akibat ketidakseimbangan distribusi data. Temuan ini menunjukkan bahwa meskipun model transformer unggul dalam performa, kualitas dan keseimbangan dataset tetap menjadi faktor penting dalam analisis sentimen.

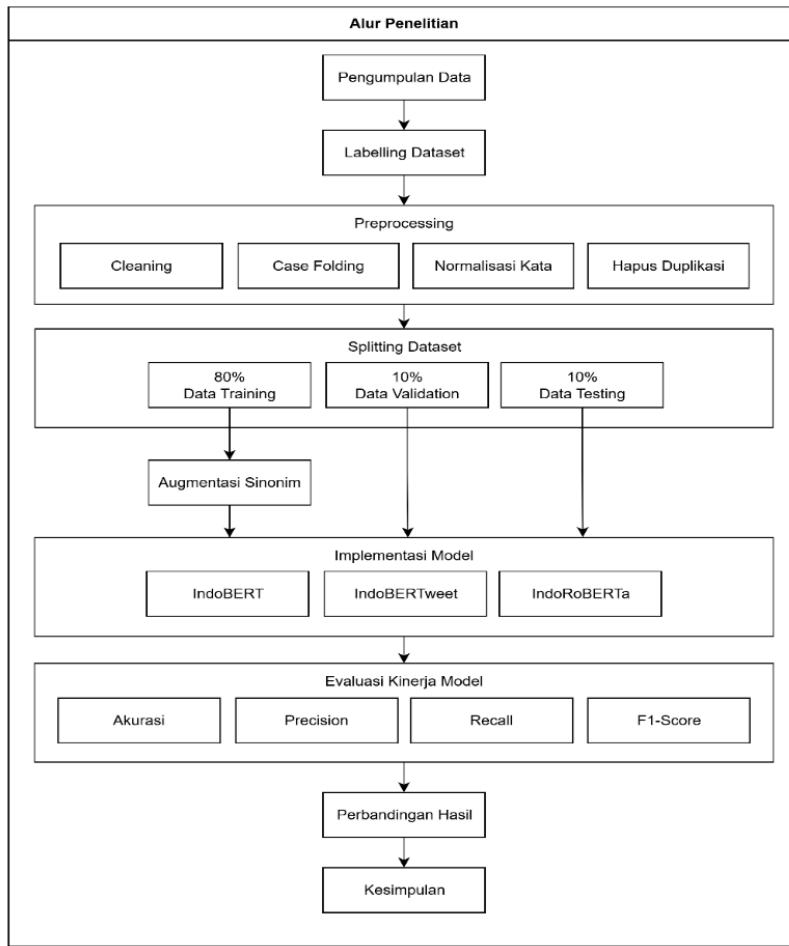
Selain IndoBERT, model transformer lain seperti RoBERTa juga menunjukkan performa yang menjanjikan. Sumihar menggunakan Indonesian RoBERTa Base Sentiment Classifier untuk menganalisis sentimen masyarakat terhadap ide calon presiden Indonesia melalui komentar di YouTube [9]. Model ini dilatih menggunakan dataset Bahasa Indonesia, termasuk SmSA, dan menghasilkan akurasi rata-rata di atas 90% pada seluruh kelas sentimen. Hasil ini mengindikasikan bahwa RoBERTa mampu memberikan representasi teks yang lebih robust dibandingkan model klasik.

Setiawan membandingkan performa model LSTM dan IndoBERTweet dalam analisis sentimen ulasan aplikasi TikTok di Google Play Store [10]. Dengan jumlah data sebanyak 22.424 ulasan, IndoBERTweet menunjukkan akurasi sebesar 80%, lebih tinggi dibandingkan LSTM yang hanya mencapai 78%. Keunggulan IndoBERTweet dikaitkan dengan proses pretraining yang secara khusus menggunakan data Twitter, sehingga lebih adaptif terhadap karakteristik bahasa informal dan singkatan yang umum digunakan di media sosial.

Penelitian yang secara spesifik membahas sentimen terhadap sistem perpajakan digital Coretax dilakukan oleh Ansori menggunakan algoritma Naïve Bayes [4]. Penelitian ini menganalisis 899 tweet dan memperoleh akurasi sebesar 77%, dengan dominasi sentimen netral. Namun, model mengalami kesulitan dalam mengenali sentimen positif akibat ketidakseimbangan data dan keterbatasan metode dalam menangkap makna semantik. Keterbatasan ini menunjukkan bahwa pendekatan konvensional belum cukup optimal untuk menganalisis opini publik terhadap layanan digital perpajakan yang memiliki konteks teknis dan kompleks.

Berdasarkan tinjauan terhadap penelitian-penelitian terdahulu tersebut, terlihat bahwa pendekatan berbasis transformer monolingual bahasa Indonesia seperti IndoBERT, IndoBERTweet, dan IndoRoBERTa memiliki keunggulan signifikan dibandingkan metode konvensional. Namun, hingga saat ini belum banyak penelitian yang secara langsung membandingkan kinerja ketiga model tersebut dalam menganalisis sentimen publik terhadap website Coretax di platform X, sehingga penelitian ini dilakukan untuk mengisi celah tersebut.

METODE PENELITIAN



GAMBAR 1. Alur Penelitian

Alur penelitian dalam studi ini dirancang secara sistematis yang dimulai dari tahapan pengumpulan data hingga penarikan kesimpulan, sebagaimana diilustrasikan pada Gambar 1. Tahap awal penelitian dimulai dengan pengumpulan data menggunakan *library* Tweet-Harvest pada platform X dengan kata kunci "coretax" untuk mengumpulkan cuitan berbahasa Indonesia yang mengandung opini terkait topik tersebut. Data yang berhasil dikumpulkan dalam format .csv kemudian melalui tahap pelabelan (*labelling dataset*) dengan metode semi otomatis, yaitu kombinasi antara penggunaan pustaka *HuggingFace* serta pelabelan manual. Pelabelan dibagi ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral.

Untuk memastikan kualitas data sebelum masuk ke tahap pemodelan, dilakukan proses *preprocessing* yang terdiri dari beberapa langkah berurutan. Pertama adalah *cleaning* untuk menghapus karakter yang tidak diperlukan seperti emoji, *mention*, URL, *hashtag*, huruf berulang, dan simbol. Langkah berikutnya adalah *case folding* untuk menyeragamkan seluruh teks menjadi huruf kecil, diikuti dengan normalisasi kata guna mengubah singkatan atau bahasa tidak baku menjadi bentuk baku. Tahap *preprocessing* diakhiri dengan penghapusan duplikasi data untuk membuang entitas yang identik setelah melalui proses pembersihan, sehingga memastikan setiap data yang digunakan bersifat unik.

Setelah data bersih, dataset dibagi (*splitting*) menjadi tiga bagian independen dengan rasio 80% untuk data latih (*training data*), 10% data validasi (*validation data*), dan 10% data uji (*testing data*). Khusus pada subset data latih, diterapkan teknik augmentasi sinonim untuk memperkaya variasi dan menambah jumlah data pada kelas minoritas. Tahapan selanjutnya adalah implementasi model menggunakan tiga arsitektur *transformer monolingual*, yaitu

IndoBERT, IndoBERTweet, dan IndoRoBERTa. Setiap model dilatih menggunakan data latih yang telah diaugmentasi dan divalidasi secara berkala menggunakan data validasi. Setelah pelatihan selesai, kinerja masing-masing model dievaluasi menggunakan data uji melalui empat metrik utama, yakni akurasi, *precision*, *recall*, dan *F1-score*. Hasil dari evaluasi ketiga model tersebut kemudian dibandingkan secara mendalam untuk mengidentifikasi arsitektur yang paling optimal dalam melakukan analisis sentimen terhadap website coretax, yang diakhiri dengan penarikan kesimpulan berdasarkan temuan utama penelitian.

HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan library Tweet-Harvest, sebuah pustaka Python yang memanfaatkan API dari platform media sosial X. Data yang dikumpulkan berupa tweet berbahasa Indonesia yang mengandung opini masyarakat terhadap website Coretax yang dikelola oleh Direktorat Jenderal Pajak.

Proses crawling dilakukan dengan menggunakan kata kunci utama “coretax”. Mekanisme crawling data dari Tweet-Harvest adalah dengan melakukan searching berdasarkan kata kunci kemudian akan dilakukan scrolling otomatis halaman pencarian untuk menjaring data tweet yang relevan. Data yang dikumpulkan disimpan dalam format CSV dan terdiri dari atribut atau kolom yang ditunjukkan pada Tabel 1. berikut.

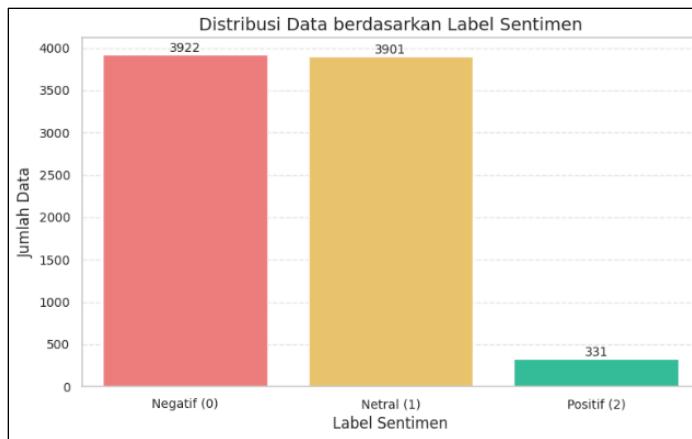
TABEL 1. Atribut Atribut dari Dataset Hasil Crawling

No.	Atribut	Deskripsi
1	conversation_id_str	ID unik dari sebuah percakapan atau thread
2	created_at	Tanggal dan waktu kapan tweet tersebut diunggah
3	favorite_count	Jumlah 'likes' atau 'favorit' yang diterima
4	full_text	Isi lengkap dari tweet yang diunggah
5	id_str	ID unik yang diberikan oleh X untuk setiap tweet
6	tweet_url	Tautan lengkap untuk mengakses tweet
7	username	Nama pengguna X yang memposting tweet
8	lang	Bahasa yang digunakan dalam tweet
9	user_id_str	ID unik dari pengguna X yang mengunggah tweet
10	quote_count	Jumlah tweet di-quote oleh pengguna lain

Dari Tabel 1. di atas menampilkan 10 atribut atau kolom dari dataset hasil *crawling* serta deskripsi dari masing-masing atribut tersebut. Atribut yang dicatat meliputi *conversation_id_str*, *created_at*, *favorite_count*, *full_text*, *id_str*, *tweet_url*, *username*, *lang*, *user_id_str*, *quote_count*, dan *reply_count*. Meskipun dataset yang dihasilkan mencakup berbagai metadata pendukung seperti parameter waktu (*created_at*), identitas unik (*id_str*, *user_id_str*), hingga metrik interaksi (*favorite_count*, *quote_count*) yang berfungsi untuk menjamin validitas dan ketertelusuran sumber data penelitian ini secara khusus hanya akan menggunakan satu atribut utama untuk tahap pemrosesan selanjutnya, yaitu *full_text*. Fokus tunggal pada atribut *full_text* ini dilakukan karena tujuan utama penelitian adalah melakukan analisis sentimen berbasis teks. Total data yang berhasil dikumpulkan sebanyak 8.154 tweet, yang diperoleh dalam rentang waktu 18 Maret hingga 31 Mei 2025. Jumlah data dan cakupan waktu tersebut disesuaikan dengan hasil *crawling* yang dapat dijangkau secara maksimal oleh library *Tweet-Harvest*, mengingat adanya batasan teknis dan keterbatasan akses historis dari API Media Sosial X.

2. Labelling Dataset

Karena data hasil crawling tidak memiliki label sentimen, dilakukan proses pelabelan menjadi tiga kelas: negatif (0), netral (1), dan positif (2). Pelabelan dilakukan dengan pendekatan semi otomatis, menggunakan model *mdhugol/indonesia-bert-sentiment-classification* dari pustaka *HuggingFace*. Setelah proses pelabelan otomatis dilakukan, masih ditemukan sejumlah data yang mengalami kesalahan dalam penentuan kategori sentimen, sehingga penulis perlu melakukan pelabelan manual menggunakan Microsoft Excel. Hasil dari pelabelan dataset tersebut dapat dilihat pada gambar 2 berikut.



GAMBAR 2. Hasil Labelling Dataset

Gambar 2. menyajikan representasi visual dari distribusi dataset yang telah melalui proses pelabelan akhir, memberikan gambaran komprehensif mengenai persepsi pengguna platform X terhadap sistem *coretax*. Berdasarkan grafik tersebut, terlihat dominasi yang signifikan pada sentimen negatif dan netral dibandingkan dengan sentimen positif. Secara rinci, kategori sentimen negatif menempati urutan tertinggi dengan jumlah 3.922 data, yang merefleksikan banyaknya keluhan, kritik, atau hambatan teknis yang disampaikan oleh masyarakat terkait implementasi *website* tersebut. Angka ini terpaut sangat tipis dengan kategori sentimen netral yang berjumlah 3.901 data, yang umumnya berisi cuitan berupa penyebaran informasi objektif, pertanyaan teknis tanpa muatan emosional, atau diskusi mengenai prosedur perpajakan tanpa memberikan penilaian subjektif. Kontras yang sangat mencolok terlihat pada kategori sentimen positif yang hanya berjumlah 331 data, menunjukkan bahwa apresiasi atau pengalaman memuaskan terhadap sistem *coretax* masih menjadi opini minoritas di media sosial. Kondisi distribusi data ini menunjukkan adanya ketidakseimbangan kelas (*imbalanced dataset*) yang sangat tajam, di mana jumlah data positif tidak sampai sepuluh persen dari total populasi data negatif maupun netral. Hal inilah yang menjadi justifikasi utama dilakukannya tahap augmentasi sinonim pada proses selanjutnya, guna memperkaya variasi data pada label positif agar model tidak cenderung hanya memprediksi kelas mayoritas (*bias*). Selain itu, dominasi sentimen negatif dan netral mengindikasikan bahwa topik "coretax" di platform X lebih banyak berfungsi sebagai ruang untuk menyampaikan aspirasi kritis dan kebutuhan akan edukasi teknis.

3. Preprocessing

Tahap preprocessing dilakukan untuk membersihkan dan menyeragamkan teks pada kolom *full_text*. Proses ini meliputi:

- a. **Cleaning**, yaitu menghapus elemen tidak relevan seperti URL, mention, hashtag, simbol, emoji, dan karakter non-ASCII. Hasilnya ditunjukkan pada Tabel 2. berikut.

TABEL 2. Hasil Cleaning

Sebelum Cleaning	Sesudah Cleaning
<p>@fefesantosa I see I see tapi keknya skrg pajak dibiarin karena CoreTax aplikasi terbaik ga sih cek penjelasan lengkapnya di artikel Suket PP 55: Tidak Muncul di Coretax</p> <p>https://t.co/9L91MWTpcD</p> <p>Suket PP 55 sudah terbit tapi fasilitasnya gak muncul di #Coretax pemotong? Minta lawan transaksi (WP UMKM) cek Fasilitas Aktif di menu Profil Saya di Coretax mereka. Kalau gak ada harus konfirmasi ulang ke KPP!</p>	<p>I see I see tapi keknya skrg pajak dibiarin karena CoreTax aplikasi terbaik ga sih cek penjelasan lengkapnya di artikel Suket PP 55 Tidak Muncul di Coretax</p> <p>Suket PP 55 sudah terbit tapi fasilitasnya gak muncul di pemotong Minta lawan transaksi WP UMKM cek Fasilitas Aktif di menu Profil Saya di Coretax mereka Kalau gak ada harus konfirmasi ulang ke KPP!</p>

- b. **Case Folding**, yaitu mengubah seluruh teks menjadi huruf kecil agar kata dengan perbedaan kapitalisasi dianggap sama. Contohnya dapat dilihat pada Tabel 3. berikut.

TABEL 3. Hasil Case Folding

Sebelum Case Folding	Sesudah Case Folding
I see I see tapi keknya skrg pajak dibiarin karena CoreTax aplikasi terbaik ga sih	i see i see tapi keknya skrg pajak dibiarin karena coretax aplikasi terbaik ga sih
cek penjelasan lengkapnya di artikel Suket PP 55 Tidak Muncul di Coretax	cek penjelasan lengkapnya di artikel suket pp 55 tidak muncul di coretax
Suket PP 55 sudah terbit tapi fasilitasnya gak muncul di pemotong Minta lawan transaksi WP UMKM cek Fasilitas Aktif di menu Profil Saya di Coretax mereka Kalau gak ada harus konfirmasi ulang ke KPP	suket pp 55 sudah terbit tapi fasilitasnya gak muncul di pemotong minta lawan transaksi wp umkm cek fasilitas aktif di menu profil saya di coretax mereka kalau gak ada harus konfirmasi ulang ke kpp

- c. **Normalisasi Kata**, yaitu mengganti kata tidak baku, singkatan, dan slang menjadi bentuk baku berdasarkan *colloquial-indonesian-lexicon*. Hasil penerapannya pada data dapat dilihat pada Tabel 4 berikut.

TABEL 4. Hasil Normalisasi Kata

Sebelum Normalisasi Kata	Sesudah Normalisasi Kata
i see i see tapi keknya skrg pajak dibiarin karena coretax aplikasi terbaik ga sih	i see i see seperti sekarang pajak dibiarin karena coretax aplikasi terbaik tidak sih
inget nder skrg pake coretax yang bikin emosi jiwa gpp dimasukin aja namanya juga pengalaman	ingat nder sekarang pakai coretax yang bikin emosi jiwa tidak apa-apa dimasukin saja namanya juga pengalaman
suket pp 55 sudah terbit tapi fasilitasnya gak muncul di pemotong minta lawan transaksi wp umkm cek fasilitas aktif di menu profil saya di coretax mereka kalau gak ada harus konfirmasi ulang ke kpp	suket pp 55 sudah terbit tapi fasilitasnya tidak muncul di pemotong minta lawan transaksi wp umkm cek fasilitas aktif di menu profil saya di coretax mereka kalau tidak ada harus konfirmasi ulang ke kpp

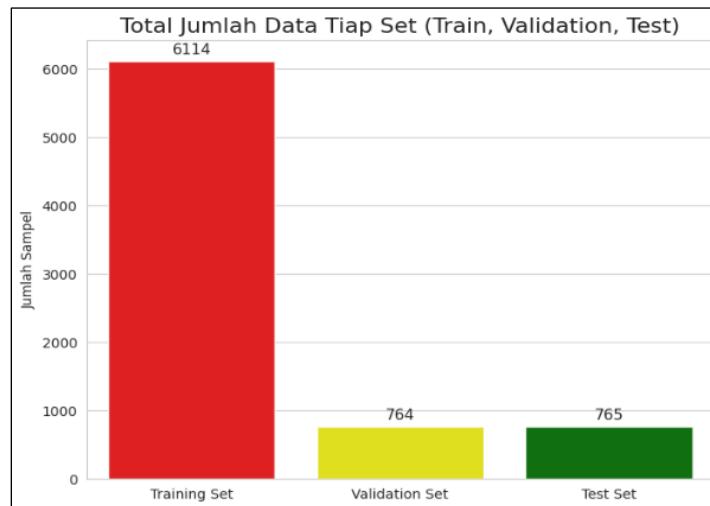
- d. **Menghapus Duplikasi**, yaitu menghapus data identik setelah pembersihan menggunakan fungsi `drop_duplicates()` dari Pandas. Hasilnya terlihat pada Tabel 4.7 berikut.

TABEL 5. Hasil Penghapusan Duplikasi Data

	Sebelum	Sesudah
Jumlah data duplikat	511	0
Data negatif	3922	3697
Data Positif	3901	3661
Data Netral	331	285

4. Split Dataset

Setelah melalui tahap preprocessing, dataset dibagi menjadi tiga subset: training (80%), validation (10%), dan testing (10%) menggunakan metode *stratified sampling* dengan fungsi `train_test_split` dari Scikit-learn. Teknik stratifikasi dipilih agar distribusi kelas sentimen tetap proporsional di setiap subset, sehingga evaluasi model lebih objektif dan bebas bias. Hasil Split Dataset dapat dilihat pada Gambar 3. berikut.

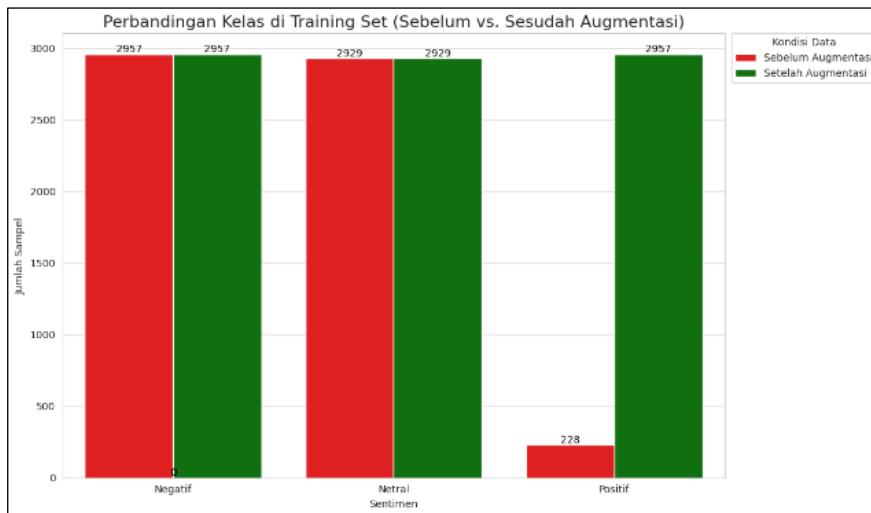


GAMBAR 3. Total Jumlah Data Tiap Set

Gambar 3 mengilustrasikan hasil pembagian dataset secara sistematis ke dalam tiga subset utama, yaitu *training set* sebanyak 6.114 data, *validation set* sebanyak 764 data, dan *testing set* sebanyak 765 data. Pembagian ini secara konsisten menerapkan rasio 80:10:10 yang dirancang untuk memberikan volume data yang cukup bagi model dalam mempelajari pola bahasa, sekaligus menyediakan dataset independen untuk pengujian kinerja. Melalui teknik *stratified sampling*, distribusi label sentimen negatif, netral, dan positif dipertahankan secara proporsional pada setiap subset guna menjamin evaluasi yang objektif dan bebas bias. Setelah tahap pembagian data ini diselesaikan, proses penelitian akan dilanjutkan ke tahap augmentasi sinonim yang secara khusus diaplikasikan pada subset data latih (*training data*). Langkah ini menjadi krusial untuk memperkaya variasi linguistik dan menambah jumlah sampel pada kelas minoritas, terutama sentimen positif yang memiliki jumlah paling sedikit, sebelum data tersebut diumpulkan ke dalam model IndoBERT, IndoBERTweet, dan IndoRoBERTa. Dengan demikian, model diharapkan memiliki sensitivitas yang lebih baik dalam mengenali nuansa opini masyarakat terhadap platform *coretax* di platform X.

5. Augmentasi Sinonim

Ketidakseimbangan kelas pada training set teridentifikasi dengan jumlah data positif (228) jauh lebih sedikit dibandingkan negatif (2.957) dan netral (2.929). Kondisi ini berpotensi membuat model bias terhadap kelas mayoritas. Untuk mengatasinya, diterapkan augmentasi sinonim menggunakan *Open Multilingual WordNet* melalui pustaka NLTK. Proses augmentasi hanya dilakukan pada training set guna mencegah *data leakage* pada set validasi dan pengujian. Hasilnya, jumlah data positif meningkat menjadi 2.957, sehingga seimbang dengan kelas negatif. Distribusi setelah augmentasi ditunjukkan pada Gambar 4. berikut.



GAMBAR 4. Perbandingan Distribusi Data Training Set Sebelum dan Sesudah Augmentasi

Gambar 4 mengilustrasikan transformasi signifikan pada komposisi *training set* melalui penerapan teknik augmentasi sinonim, yang bertujuan untuk menyeimbangkan distribusi kelas sentimen sebelum tahap pelatihan model dimulai. Pada kondisi awal, sebagaimana terlihat pada sisi distribusi sebelum augmentasi, terdapat ketimpangan data yang sangat ekstrem di mana kelas positif hanya berjumlah 228 sampel, berbanding terbalik dengan kelas negatif yang mencapai 2.957 sampel dan kelas netral sebanyak 2.929 sampel. Ketidakseimbangan kelas ini merupakan tantangan kritis dalam pengembangan model klasifikasi, karena algoritma *transformer* cenderung akan mengalami bias terhadap kelas mayoritas dan kesulitan dalam mempelajari fitur unik dari kelas minoritas. Melalui prosedur augmentasi menggunakan *Open Multilingual WordNet*, variasi linguistik pada kelas positif ditingkatkan secara sistematis hingga jumlahnya mencapai 2.957 sampel, sehingga kini memiliki bobot yang setara dengan kelas negatif.

6. Implementasi Model

Tahap ini melakukan fine-tuning pada tiga model transformer monolingual berbahasa Indonesia: IndoBERT, IndoBERTweet, dan IndoRoBERTa. Proses dilakukan menggunakan pustaka *Hugging Face Transformers* dengan dataset Coretax yang telah melalui preprocessing dan pelabelan. Tokenisasi menggunakan tokenizer bawaan tiap model. Pelatihan dilakukan menggunakan objek Trainer, yang mengintegrasikan model, data latih/validasi, serta evaluasi metrik melalui *compute metrics*. Strategi early stopping diterapkan untuk mencegah *overfitting* dan *underfitting*. Parameter pelatihan ditunjukkan pada Gambar 5.

	Model	Tokenizer	Epoch	Batch Size	Early Stopping
0	IndoBERT	indobenchmark/indobert-base-p1	10	16	Patience = 2
1	IndoBERTweet	indolem/indobertweet-base-uncased	10	16	Patience = 2
2	IndoRoBERTa	flax-community/indonesian-roberta-base	10	16	Patience = 2

GAMBAR 5. Konfigurasi Hyperparameter untuk Fine Tuning Model

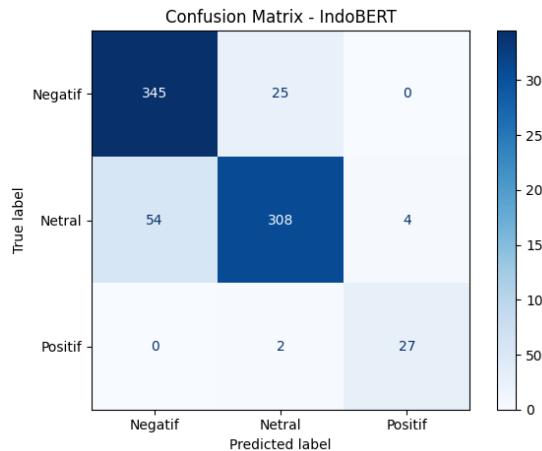
Gambar 5 merinci konfigurasi *hyperparameter* yang krusial bagi proses *fine-tuning* tiga model *transformer monolingual*, yaitu IndoBERT, IndoBERTweet, dan IndoRoBERTa. Setiap arsitektur menggunakan *tokenizer* bawaan yang spesifik, seperti indolem/indobertweet-base-uncased untuk IndoBERTweet dan flax-community/indonesian-roberta-base untuk IndoRoBERTa, guna memastikan pemrosesan teks yang optimal. Seluruh pelatihan dilakukan selama 10 *epoch* dengan ukuran *batch* (*batch size*) sebesar 16 untuk menyeimbangkan kecepatan komputasi dan stabilitas optimasi. Penerapan parameter yang seragam ini bertujuan menciptakan lingkungan eksperimen yang adil, sehingga perbedaan hasil evaluasi nantinya benar-benar merepresentasikan keunggulan masing-masing arsitektur secara objektif. Selain itu, strategi mitigasi berupa *Early Stopping* dengan nilai *patience* 2 diimplementasikan untuk

mencegah terjadinya *overfitting*. Mekanisme ini secara otomatis menghentikan pelatihan jika performa pada data validasi tidak menunjukkan perbaikan signifikan dalam dua iterasi berturut-turut. Dengan kendali *hyperparameter* yang sistematis ini, model diharapkan mampu menghasilkan klasifikasi sentimen yang akurat terhadap implementasi sistem coretex.

7. Evaluasi Kinerja Model

Evaluasi dilakukan terhadap IndoBERT, IndoBERTweet, dan IndoRoBERTa menggunakan data uji dengan metrik akurasi, precision, recall, f1-score, serta confusion matrix.

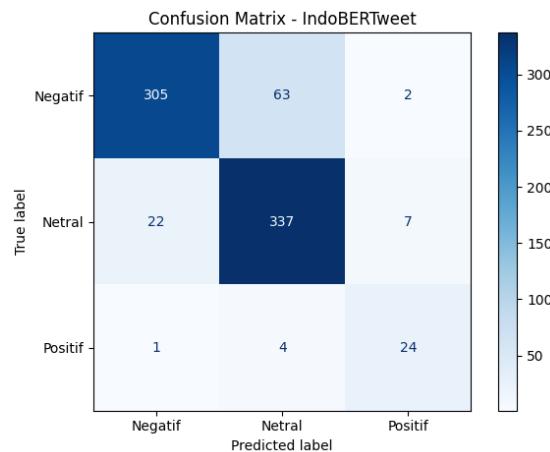
a. IndoBERT



GAMBAR 6. Hasil Confusion Matrix IndoBERT

Gambar 6 menunjukkan bahwa IndoBERT berhasil mengklasifikasikan 345 data Negatif dengan benar dari total 370, sedangkan 25 data salah diprediksi sebagai Netral. Untuk kelas Netral, sebanyak 308 data dari 366 tepat terkласifikasi, sementara 54 salah diprediksi sebagai Negatif dan 4 sebagai Positif. Pada kelas Positif, model mengenali 27 data dengan benar dari 29 yang tersedia. Secara umum, IndoBERT unggul dalam mengenali kelas Negatif, namun masih sering keliru membedakan Netral dengan Negatif.

b. IndoBERTweet

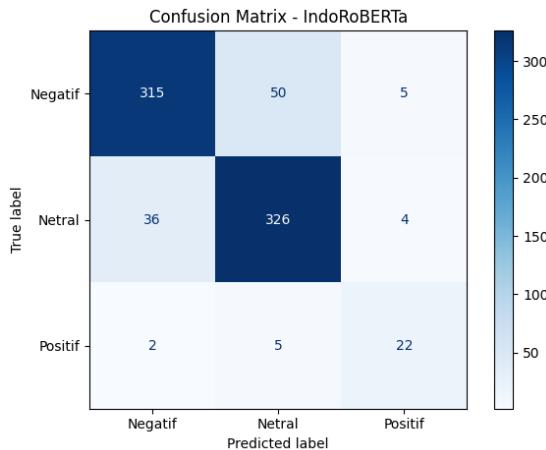


GAMBAR 7. Hasil Confusion Matrix IndoBERTweet

Gambar 7 memperlihatkan IndoBERTweet mampu mengklasifikasikan 337 data Netral dengan benar dari 366, dengan kesalahan sebanyak 22 data yang diprediksi sebagai Negatif. Pada kelas Negatif, 305 data dari 370 berhasil terkласifikasi dengan tepat, tetapi terdapat 63 data yang salah diprediksi sebagai Netral. Untuk kelas Positif, model mengenali 24 data dengan benar dari 29. Model ini menunjukkan keungulan pada kelas Netral,

meskipun kelemahan utama terletak pada pembedaan kelas Negatif yang sering salah dikategorikan sebagai Netral.

c. IndoRoBERTa



GAMBAR 8. Hasil Confusion Matrix IndoRoBERTa

Gambar 8 menampilkan IndoRoBERTa dengan hasil yang relatif seimbang. Pada kelas Negatif, sebanyak 315 data dari 370 diklasifikasikan dengan benar, sementara 50 data salah diprediksi sebagai Netral. Untuk kelas Netral, 326 data dari 366 berhasil diklasifikasikan dengan tepat, namun 36 data keliru dianggap sebagai Negatif. Pada kelas Positif, 22 data dari 29 terklasifikasi benar, sedangkan 5 salah diprediksi sebagai Netral dan 2 sebagai Negatif. Model ini cukup konsisten pada kelas mayoritas, tetapi kesulitan dalam mengenali kelas Positif.

8. Perbandingan Hasil

Perbandingan kinerja IndoBERT, IndoBERTweet, dan IndoRoBERTa dilakukan menggunakan classification report yang mencakup precision, recall, f1-score, dan accuracy. Hasil macro average disajikan pada Tabel 3.6 berikut

TABEL 7. Hasil Classification Report Macro Average

Model	Precision	Recall	F1-Score	Accuracy
IndoBERT	0.8850	0.9017	0.8920	0.8889
IndoBERTweet	0.8721	0.8781	0.8735	0.8850
IndoRoBERTa	0.7983	0.8804	0.8308	0.8732

Tabel 7 menunjukkan bahwa IndoBERT memperoleh hasil terbaik dengan precision 0.8850, recall 0.9017, f1-score 0.8920, dan akurasi 0.8889. Nilai ini mendapatkan keseimbangan yang baik antara ketepatan dan kemampuan mendekripsi data yang relevan. IndoBERTweet berada di posisi kedua dengan f1-score 0.8735 dan akurasi 0.8850, performanya cukup dekat dengan IndoBERT meskipun sedikit lebih rendah. IndoRoBERTa menempati posisi terendah dengan f1-score 0.8308 dan akurasi 0.8732; model ini memiliki recall tinggi (0.8804), tetapi precision rendah (0.7983) sehingga prediksinya kurang presisi. Secara keseluruhan, IndoBERT merupakan model dengan performa paling optimal, diikuti IndoBERTweet, sedangkan IndoRoBERTa relatif kurang akurat dan presisi.

KESIMPULAN

Penelitian ini berhasil menjawab rumusan masalah mengenai perbandingan efektivitas model *transformer monolingual* dalam melakukan analisis sentimen terhadap implementasi *website coretax* pada platform X. Melalui

proses pengumpulan data menggunakan *library* Tweet-Harvest, didapatkan dataset sebanyak 8.154 cuitan yang menunjukkan tingginya attensi publik terhadap topik perpajakan ini. Setelah melalui tahap pembersihan data, ditemukan bahwa persepsi masyarakat didominasi oleh sentimen negatif (3.922 data) dan netral (3.901 data), sementara sentimen positif hanya mewakili porsi kecil (331 data). Temuan ini mengindikasikan bahwa sebagian besar opini pengguna di platform X lebih banyak berisi kritik teknis atau diskusi prosedural tanpa muatan emosional positif yang signifikan.

Secara metodologis, penelitian ini membuktikan bahwa penanganan ketidakseimbangan data (*imbalanced dataset*) melalui teknik augmentasi sinonim pada *training set* merupakan langkah yang krusial. Dengan meningkatkan jumlah sampel kelas positif dari 228 menjadi 2.957, model mampu mempelajari fitur-fitur sentimen minoritas secara lebih adil. Hal ini tercermin dari keberhasilan proses *fine-tuning* pada tiga model yang diuji menggunakan parameter terkendali, seperti penggunaan 10 *epoch* dan strategi *early stopping* untuk mencegah degradasi performa model.

Hasil akhir evaluasi menunjukkan bahwa model IndoBERT (indobenchmark/indobert-base-p1) tampil sebagai arsitektur yang paling optimal dan konsisten dibandingkan model lainnya. IndoBERT berhasil mencapai skor metrik tertinggi dengan nilai *precision* 0.8850, *recall* 0.9017, *f1-score* 0.8920, serta akurasi sebesar 0.8889. Performa ini mengungguli IndoBERTweet yang berada di posisi kedua dan IndoRoBERTa yang menunjukkan hasil terendah dalam eksperimen ini. Tingginya nilai *recall* dan *f1-score* pada IndoBERT membuktikan bahwa model tersebut sangat sensitif dan akurat dalam mengenali nuansa bahasa Indonesia yang digunakan masyarakat saat membicarakan *coretax*. Oleh karena itu, IndoBERT sangat direkomendasikan untuk digunakan dalam sistem pemantauan opini publik otomatis guna membantu instansi terkait dalam mengevaluasi efektivitas sistem perpajakan digital mereka berdasarkan aspirasi masyarakat secara *real-time*.

TINJAUAN PUSTAKA

- [1] A. Sitanggang, Y. Umaidah, Y. Umaidah, R. I. Adam, and R. I. Adam, “ANALISIS SENTIMEN MASYARAKAT TERHADAP PROGRAM MAKAN SIANG GRATIS PADA MEDIA SOSIAL X MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4902.
- [2] W. Maria Sonia Nadiah Hutagalung and N. Jaya Perdana, “ANALISIS SENTIMEN PADA OPINI KENAIKAN HARGA BAHAN BAKAR MINYAK PADA MEDIA SOSIAL TWITTER,” *Jurnal Sistem Informasi Dan Teknik Komputer*, vol. 8, no. 2, 2023.
- [3] N. Fadhilah Az-Haari, D. Juardi, and A. Jamaludin, “ANALISIS SENTIMEN TERHADAP BOIKOT BRAND PRO-ISRAEL PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES (STUDI KASUS: STARBUCKS),” 2024.
- [4] A. Faradhisna Ansori, I. Nailah Ramadhani, C. Rahmi Anissa, and S. Amelia Putri, “ANALISIS SENTIMEN MASYARAKAT INDONESIA DI TWITTER TERHADAP SISTEM PERPAJAKAN ‘CORETAX’ MENGGUNAKAN METODE NAÏVE BAYES,” 2025.
- [5] R. Jevanya Meliala, N. Indah Chasanah, J. Steven Rajali Manik, T. Muhammad Pasya, H. Rubina Lestari, and T. Rekayasa Perangkat Lunak, “ANALISIS SENTIMEN TAGAR #KABURAJADULU PILIHAN MIGRASI KE JEPANG PADA PLATFORM X DENGAN NLP,” *Journal of Digital Business and Technology Innovation (DBESTI)*, vol. 2, no. 1, pp. 1–7, 2025.
- [6] D. Fatma Sjoraida, B. Wibawa, K. Guna, and D. Yudhakusuma, “Analisis Sentimen Film Dirty Vote Menggunakan BERT (Bidirectional Encoder Representations from Transformers),” *Universitas Langlangbuana*, vol. 8, no. 2, 2024, doi: 10.35870/jti.
- [7] L. Geni, E. Yulianti, and D. I. Sensuse, “Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using Bert Language Models,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 3, pp. 746–757, Aug. 2023, doi: 10.26555/jiteki.v9i3.26490.
- [8] Y. A. Singgalen, “Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data,” *Journal of Information System Research*, vol. 6, no. 2, pp. 976–986, 2025, doi: 10.47065/josh.v6i2.6505.
- [9] Y. ’ El and P. Sumihar, “Sentiment Analysis of Public Opinions Regarding ‘Ideas of Presidential Candidates’ in YouTube Video Comments with Robustly Optimized BERT Pretraining Approach,” 2024.

- [10] J. C. Setiawan, K. M. Lhaksmana, and B. Bunyamin, “Sentiment Analysis of Indonesian TikTok Review Using LSTM and IndoBERTweet Algorithm,” JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 8, no. 3, pp. 774–780, Aug. 2023, doi: 10.29100/jipi.v8i3.3911.