

Public Sentiment Analysis on Corruption Issues in Indonesia Using IndoBERT Fine-Tuning, Logistic Regression, and Linear SVM

Maria Fatima Kono^{1*}, Ika Nur Fajri^{2*}, Yoga Pristyanto^{3*}

* Sistem Informasi, Universitas Amikom Yogyakarta

mariafatimakono@students.amikom.ac.id¹, fajri@amikom.ac.id², yoga.pristyanto@amikom.ac.id³

Article Info

Article history:

Received 2025-07-28
Revised 2025-09-08
Accepted 2025-09-20

Keyword:

*Sentiment Analysis,
IndoBERT,
Fine-Tuning,
Logistic Regression,
Linear SVM,
Social Media,
Corruption,
SMOTE.*

ABSTRACT

Sentiment analysis is a method in *Natural Language Processing* (NLP) that aims to understand public perceptions based on textual data from social media. Opinions expressed in digital platforms play an important role as they reflect public trust and attitudes toward strategic issues in Indonesia. This study aims to compare the performance of three IndoBERT-based approaches for sentiment classification, namely IndoBERT with full *fine-tuning*, IndoBERT as a feature extractor combined with Logistic Regression, and IndoBERT as a feature extractor combined with Linear SVM. The dataset was collected through the Twitter API, consisting of 2,012 tweets, which after preprocessing and balancing resulted in 2,252 labeled data for positive and negative sentiments. The preprocessing stage included cleansing, normalization, tokenization, stopword removal, and stemming. The dataset was then split into 80% training data, 10% validation data, and 10% testing data. Experimental results show that IndoBERT with full *fine-tuning* achieved the best performance, with an accuracy of 82.67%, an F1-score of 82.35%, and an AUC value of 0.87. Logistic Regression and Linear SVM produced lower accuracies of 80.20% and 78.22%, respectively. These findings indicate that *fine-tuned* IndoBERT is more effective in capturing the semantic nuances of the Indonesian language, while the *non fine-tuning* approaches offer better computational efficiency at the cost of reduced accuracy. This study contributes to the development of NLP methods for the Indonesian language, particularly in sentiment analysis, and highlights the potential of transformer-based models for analyzing strategic issues in social media.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Korupsi merupakan salah satu persoalan serius di Indonesia yang berdampak luas terhadap perekonomian, tata kelola pemerintahan, serta tingkat kepercayaan publik terhadap institusi negara[1]. Berdasarkan laporan Transparency International (2024), Indeks Persepsi Korupsi (IPK) Indonesia tercatat pada skor 34 dari 100 dan menempati peringkat ke-115 dari 180 negara di dunia. Kondisi ini menunjukkan bahwa praktik korupsi masih menjadi masalah utama, termasuk di lingkungan badan usaha milik negara (BUMN) yang memiliki peranan strategis dalam menopang perekonomian nasional[2].

Dalam era digital, media sosial menjadi ruang utama masyarakat untuk mengekspresikan opini dan sikap terhadap

isu-isu strategis, termasuk dugaan korupsi yang melibatkan institusi publik. Opini publik yang tercermin melalui media sosial mencakup beragam persepsi, mulai dari kritik terhadap tata kelola hingga dukungan atas upaya pemberantasan korupsi. Analisis terhadap opini publik tersebut dapat dilakukan melalui *sentiment analysis* atau *opinion mining*, yaitu teknik *text mining* untuk mengklasifikasikan opini menjadi kategori positif atau negatif dengan memanfaatkan *Natural Language Processing* (NLP) dan algoritma pembelajaran mesin [3], [4], [5].

Beberapa penelitian sebelumnya telah mengkaji isu korupsi dan opini publik di Indonesia menggunakan algoritma pembelajaran mesin klasik. Misalnya, penelitian dengan Support Vector Machine (SVM) menunjukkan akurasi 82% pada isu korupsi[6], sementara Logistic Regression dengan

representasi TF-IDF mencapai akurasi 89% dan F1-score 86,5%[3]. Linear SVM juga dilaporkan menghasilkan akurasi 90,17% dengan F1-score 0,902[7]. Selain itu, metode berbasis *deep learning* seperti LSTM juga mulai diterapkan untuk analisis sentimen[5]. Walaupun hasilnya cukup baik, sebagian besar penelitian tersebut masih terbatas pada algoritma tradisional atau arsitektur umum, bukan model *transformer* khusus bahasa Indonesia.

Perkembangan terbaru menunjukkan bahwa IndoBERT sebagai model *transformer* berbasis bahasa Indonesia mampu memberikan performa unggul dalam memahami konteks teks. Penelitian pada berbagai domain, seperti ulasan hotel, pendidikan, dan e-commerce, menunjukkan akurasi di atas 90%[8], [9], [10]. Namun, kajian yang secara sistematis membandingkan kinerja IndoBERT dengan *fine-tuning penuh* dan pendekatan *non fine-tuning* (menggunakan IndoBERT sebagai ekstraktor fitur yang kemudian diklasifikasikan dengan Logistic Regression atau SVM) pada isu-isu strategis di Indonesia masih sangat terbatas. Kondisi ini menunjukkan adanya celah penelitian yang penting untuk diisi, sehingga membuka peluang bagi penelitian ini untuk memberikan kontribusi akademis maupun praktis.

Berdasarkan uraian tersebut, penelitian ini memiliki urgensi ganda. Dari sisi akademis, penelitian ini berkontribusi dalam pengembangan metode analisis sentimen berbahasa Indonesia melalui perbandingan kinerja IndoBERT *fine-tuning* dan *non fine-tuning*. Dari sisi praktis, hasil penelitian diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai persepsi publik terhadap isu strategis serta menjadi acuan dalam merancang strategi komunikasi publik yang lebih efektif. Oleh karena itu, penelitian ini bertujuan untuk (1) mengumpulkan opini publik terkait isu strategis yang berkembang di media sosial, (2) melakukan klasifikasi sentimen dengan tiga pendekatan utama, yaitu IndoBERT dengan *fine-tuning penuh*, IndoBERT tanpa *fine-tuning* dengan Logistic Regression, serta IndoBERT tanpa *fine-tuning* dengan Linear SVM, dan (3) membandingkan kinerja masing-masing pendekatan berdasarkan akurasi, presisi, recall, F1-score, serta efisiensi komputasi.

II. METODE

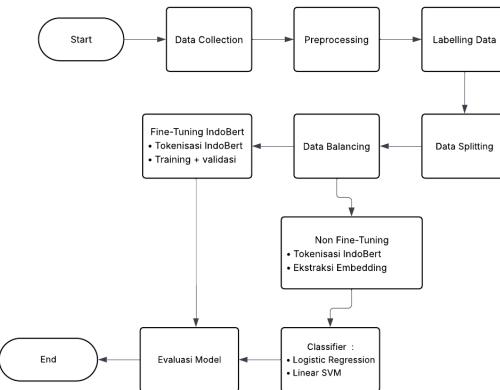
Penelitian ini menggunakan metode supervised learning berbasis klasifikasi teks dalam ranah penelitian kuantitatif. Fokus utama penelitian adalah penerapan IndoBERT dengan *fine-tuning* penuh untuk membangun model analisis sentimen. Proses penelitian dilakukan secara berurutan mulai dari pengumpulan data, pra-pemrosesan, pelatihan model, hingga evaluasi performa. Pada pendekatan ini, IndoBERT dilatih ulang secara end-to-end agar dapat menyesuaikan bobot model terhadap dataset yang digunakan. dengan representasi teks melalui tokenisasi sub-kata [3], [7].

Sebagai banding, penelitian ini juga mengevaluasi kinerja IndoBERT tanpa *fine-tuning*. Pada jalur ini, embedding IndoBERT diekstraksi dan kemudian digunakan sebagai fitur masukan untuk dua algoritma klasifikasi

tradisional, yaitu Logistic Regression dan Linear SVM [11], [12], [13]. Dengan cara ini, IndoBERT hanya berperan sebagai penyedia representasi teks, sementara proses klasifikasi dilakukan oleh algoritma di luar model transformer. Perbandingan ini bertujuan untuk menilai sejauh mana *fine-tuning* dapat meningkatkan kualitas prediksi dibandingkan pendekatan non-*fine-tuning*.

Kinerja semua model diukur menggunakan metrik akurasi, presisi, recall, dan F1-score [8], [14]. Pada jalur tanpa *fine-tuning* digunakan metode k-fold cross-validation, sedangkan jalur *fine-tuning* menerapkan skema pembagian data 80/10/10 untuk pelatihan, validasi, dan pengujian [15]. Proses pelatihan model IndoBERT dengan *fine-tuning* dilakukan menggunakan AdamW optimizer, dengan learning rate 2e-5, batch size 16, serta 3 epoch pelatihan [16]. Evaluasi tetap menggunakan metrik yang sama pada kedua jalur, sehingga hasil perbandingan dapat dianalisis secara objektif.

Diagram alir berikut akan memperlihatkan urutan langkah utama yang ditempuh dalam penelitian ini.



Gambar 1. Tahapan penelitian dengan IndoBERT fine-tuning dan non-fine-tuning

A. Data Collection

Data yang digunakan dalam penelitian ini diperoleh dari media sosial X (dahulu Twitter), yang merupakan salah satu platform komunikasi publik paling aktif dan banyak digunakan masyarakat untuk menyampaikan opini terhadap isu-isu sosial, politik, maupun korporasi. Fokus penelitian ini adalah pada persepsi publik terkait isu korupsi pada BUMN strategis [2], [3].

Pengumpulan data dilakukan menggunakan Twitter API melalui skrip Python yang dikembangkan secara mandiri. Pendekatan ini memungkinkan peneliti untuk mengekstrak tweet secara langsung sesuai dengan kriteria penelitian, sehingga kontrol terhadap proses pengambilan data dapat dilakukan dengan lebih baik. Kata kunci yang digunakan adalah “kasus Pertamina” dan “PT Pertamina”, sehingga semua tweet yang membahas isu tersebut dapat dikumpulkan secara sistematis[17].

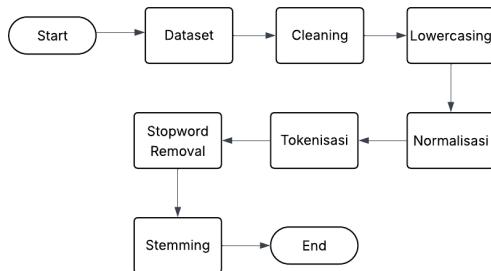
Untuk memperluas jangkauan dan mempermudah pengumpulan, sebagian data juga diambil menggunakan skrip crawling berbasis Node.js dengan paket tweet-harvest. Skrip ini mengekstrak tweet sesuai kata kunci dan menyimpannya

dalam file CSV dengan nama PT_Pertamina.csv, dengan batas maksimal pengambilan 2.000 tweet. Sebelum eksekusi skrip, lingkungan Python dan Node.js dipersiapkan, termasuk instalasi paket pandas untuk pengolahan data dan Node.js untuk menjalankan paket tweet-harvest. Autentikasi terhadap API Twitter dilakukan menggunakan Twitter Auth Token agar proses pengambilan data berjalan lancar[18], [19].

Periode pengambilan data dibatasi dari Januari hingga Mei 2025, sehingga dataset yang diperoleh merepresentasikan opini publik terkini dalam kurun waktu lima bulan. Hasil pengumpulan ini menghasilkan total 2.012 tweet, yang siap dianalisis lebih lanjut pada tahap pra-pemrosesan dan pelabelan[17]. Dengan pendekatan ini, penelitian memperoleh data mentah yang relevan, representatif, dan dapat diandalkan untuk menganalisis sentimen publik terhadap isu korupsi pada BUMN strategis [5].

B. Pre-processing

Setelah pengumpulan dataset, tahap selanjutnya adalah pra-pemrosesan atau preprocessing, yang bertujuan untuk membersihkan dan menyiapkan data agar siap dianalisis menggunakan metode klasifikasi sentimen. Tweet yang diambil dari media sosial X umumnya mengandung berbagai elemen yang tidak relevan atau mengganggu analisis, seperti mention, URL, emoji, dan karakter khusus. Oleh karena itu, preprocessing menjadi langkah penting untuk meningkatkan kualitas dataset.



Gambar 2. Tahapan preprocessing data

1) *Data Cleaning*: Tahap awal dalam pra-pemrosesan teks adalah data cleaning, yaitu membersihkan dataset dari elemen yang tidak relevan atau mengganggu. Proses ini mencakup penghapusan data duplikat, entri kosong, komentar netral, URL, angka, emoji, tanda baca, dan simbol khusus. Dengan langkah ini, data menjadi lebih bersih, bebas dari noise, dan siap dianalisis lebih lanjut [18].

2) *Lowercasing*: Setelah data dibersihkan, dilakukan lowercasing, yaitu mengubah seluruh huruf dalam teks menjadi huruf kecil. Misalnya, istilah “KORUPSI”, “Korupsi”, atau “korupsi” akan diperlakukan sama oleh model. Tahap ini membantu menyamakan representasi kata sehingga variasi huruf kapital tidak memengaruhi hasil penelitian [4].

3) *Normalisasi*: Normalisasi bertujuan menyamakan kata-kata yang tidak baku, slang, atau singkatan menjadi

bentuk baku sesuai Kamus Besar Bahasa Indonesia (KBBI) atau kamus manual dari analisis dataset. Contoh: “gak” atau “nggak” diubah menjadi “tidak”, dan “btw” diubah menjadi “ngomong-ngomong”. Normalisasi penting karena komentar publik sering menggunakan bahasa sehari-hari yang beragam, sehingga meningkatkan konsistensi data dan memudahkan representasi numerik teks [19].

4) *Tokenisasi*: Tokenisasi memecah teks menjadi unit kata atau token. Contoh: kalimat “PT Pertamina korupsi uang rakyat” diubah menjadi token [PT, Pertamina, korupsi, uang, rakyat]. Tahap ini memungkinkan setiap kata diperlakukan sebagai fitur individual dalam model klasifikasi dan memudahkan tahap berikutnya, seperti stopword removal dan pembuatan vektor kata [18].

5) *Stopword removal*: Stopword removal digunakan untuk menghapus kata-kata umum yang tidak memberikan kontribusi signifikan terhadap analisis sentimen, seperti “dan”, “yang”, atau “di”. Penghapusan ini membantu fokus pada kata-kata relevan terhadap sentimen publik tentang kasus korupsi dan mengurangi noise dalam dataset[4].

6) *Stemming*: Stemming Mengubah kata berimbuhan menjadi bentuk dasar untuk menyatukan variasi kata. Stemming mengembalikan kata ke bentuk dasarnya. Contoh: “menyalahgunakan” → “salahguna”, “penyalahgunaan” → “salahguna”. Tahap ini menyatukan variasi kata yang memiliki makna sama, menyederhanakan kompleksitas morfologi bahasa Indonesia, dan meningkatkan akurasi representasi fitur dalam model klasifikasi sentimen [5].

Dengan penerapan seluruh tahap pra-pemrosesan ini, dataset komentar publik mengenai kasus korupsi menjadi lebih bersih, konsisten, dan siap digunakan untuk membangun model klasifikasi sentimen. Proses ini penting untuk meningkatkan performa model, baik menggunakan IndoBERT dengan fine-tuning maupun IndoBERT tanpa fine-tuning dengan Logistic Regression atau Linear SVM, karena mengurangi noise, menyamakan kata, dan menyederhanakan kompleksitas bahasa publik.

C. Labeling Data

Tahap pelabelan data (labeling) dilakukan setelah pra-pemrosesan selesai. Dalam penelitian ini, pelabelan dilakukan secara otomatis menggunakan metode analisis sentimen berbasis NLP, sehingga setiap tweet dianalisis dan diklasifikasikan ke dalam dua kelas utama, yaitu positif (1) dan negatif (0), berdasarkan konteks dan ekspresi emosional yang terkandung dalam teks. Proses labeling otomatis melibatkan beberapa tahapan teknis. Pertama, setiap tweet diterjemahkan ke bahasa Inggris untuk memastikan kompatibilitas dengan model analisis sentimen. Kedua, teks dibersihkan kembali menggunakan fungsi pembersihan (*clean_tweet()*) untuk menghapus karakter, simbol, atau elemen yang tidak relevan. Ketiga, setiap tweet dianalisis

menggunakan fungsi analisis sentimen yang menghasilkan label dan skor keyakinan prediksi. Hasil dari proses ini disimpan dalam dataset sehingga setiap tweet memiliki informasi mengenai kelas sentimen dan tingkat kepercayaan prediksi. Pendekatan otomatis ini memungkinkan pelabelan dataset yang cukup besar, yakni 2.012 tweet, dilakukan secara cepat, konsisten, dan efisien dibandingkan pelabelan manual.

Label netral sengaja diabaikan dalam penelitian ini karena tweet dengan sentimen netral umumnya tidak mengandung opini yang jelas dan dapat menimbulkan ambiguitas dalam proses klasifikasi. Kehadiran data netral dapat menyulitkan model untuk membedakan pola antara sentimen positif dan negatif, sehingga dapat menurunkan akurasi prediksi. Dengan menghapus data netral, dataset menjadi lebih fokus pada dua kelas utama yang benar-benar mencerminkan opini publik terkait kasus korupsi pertamina. Strategi ini juga mendukung pelatihan model berbasis TF-IDF dengan Logistic Regression atau Linear SVM, maupun model berbasis IndoBERT, sehingga prediksi sentimen publik menjadi lebih akurat, konsisten, dan dapat diandalkan [3].

Hasil akhirnya adalah dataset yang siap digunakan untuk analisis lebih lanjut, terdiri dari dua kelas sentimen (positif dan negatif), dengan setiap tweet memiliki label dan skor keyakinan prediksi. Dataset ini bebas dari ambiguitas yang disebabkan oleh tweet netral dan siap untuk tahap berikutnya, seperti pembagian *train-validation-test* dan penerapan teknik balancing seperti SMOTE.

D. Data Plitting

Dataset yang telah bersih dan diberi label sentimen kemudian dibagi menjadi beberapa subset untuk keperluan pelatihan, validasi, dan pengujian model klasifikasi. Pembagian dataset ini bertujuan agar model dapat belajar dari data yang cukup, divalidasi secara objektif, dan dievaluasi pada data yang belum pernah dilihat sebelumnya. Dalam penelitian ini, dataset dibagi dengan rasio 80:10:10, di mana 80% data digunakan sebagai training set untuk melatih model mengenali pola dan fitur yang memengaruhi sentimen, 10% sebagai validation set untuk memantau performa model selama pelatihan dan membantu pemilihan hyperparameter, serta 10% sebagai test set untuk mengukur kinerja akhir model pada data baru sehingga hasil evaluasi lebih

Pembagian dilakukan secara stratified split, sehingga proporsi kelas positif dan negatif tetap seimbang di setiap subset. Hal ini penting karena dataset biasanya tidak seimbang, dan stratified split memastikan model belajar dari distribusi kelas yang representatif. Selain itu, untuk model berbasis TF-IDF, training set diimbangi menggunakan SMOTE (*Synthetic Minority Oversampling Technique*), yang menambahkan sampel sintetis pada kelas minoritas agar distribusi kelas menjadi seimbang. Secara teknis, proses pembagian dilakukan dalam dua langkah, yaitu pertama membagi dataset menjadi training set (80%) dan subset sementara (20%), kemudian subset sementara dibagi menjadi validation set (10%) dan test set (10%). Dengan pembagian ini, dataset siap digunakan untuk membangun dan

menevaluasi model klasifikasi sentimen, baik menggunakan IndoBERT maupun metode berbasis TF-IDF dengan Logistic Regression atau Linear SVM[10], [13].

E. Data Balancing

Distribusi kelas sentimen pada dataset penelitian ini tidak seimbang, di mana jumlah tweet dengan label negatif lebih banyak dibandingkan tweet positif. Ketidakseimbangan ini berpotensi menyebabkan model klasifikasi lebih condong memprediksi kelas mayoritas (negatif), sehingga kemampuan model dalam memprediksi kelas minoritas (positif) menjadi rendah. Untuk mengatasi masalah ini, dilakukan data balancing pada training set menggunakan *SMOTE (Synthetic Minority Oversampling Technique)*. SMOTE bekerja dengan membuat sampel sintetis baru untuk kelas minoritas melalui interpolasi antara sampel minoritas asli dan tetangga terdekatnya. Secara matematis, sampel baru dihitung dengan rumus:

$$x_{new} = x_i + \delta * (x_{zi} - x_i)$$

di mana $(x_{zi} - x_i)$ adalah sampel minoritas asli, x_i adalah salah satu tetangga terdekat dari $(x_{zi} - x_i)$ dan δ adalah bilangan acak antara 0 dan 1. Dengan cara ini, variasi data minoritas meningkat, sehingga distribusi kelas menjadi lebih seimbang [9].

Penerapan SMOTE hanya dilakukan pada training set, sedangkan validation set dan test set tetap menggunakan data asli agar evaluasi model mencerminkan kondisi nyata dan tidak bias [10], [18], [19]. Metode ini benar-benar meningkatkan kinerja pada kelas minoritas karena model memperoleh lebih banyak sampel representatif dari kelas positif, variasi fitur minoritas bertambah sehingga pola unik dari kelas minoritas dapat dipelajari dengan lebih baik, dan risiko overfitting pada kelas mayoritas dapat diminimalkan[6], [12]. Dengan distribusi yang lebih seimbang, model dapat mengenali fitur penting dari kedua kelas sehingga prediksi menjadi lebih stabil dan akurat[12], [13].

Dengan penerapan SMOTE, model berbasis TF-IDF dengan Logistic Regression atau Linear SVM maupun IndoBERT mampu meningkatkan kemampuan klasifikasi pada kedua kelas, sehingga prediksi sentimen publik terkait kasus PT Pertamina menjadi lebih konsisten, akurat, dan dapat diandalkan [6], [10], [11]. Implementasi ini juga memungkinkan pengecekan distribusi kelas sebelum dan sesudah SMOTE, sehingga kualitas data training dapat diverifikasi dengan jelas [17], [19].

F. Fine-tuning Indobert

Proses fine-tuning dilakukan dengan memanfaatkan model IndoBERT pre-trained yang kemudian disesuaikan untuk tugas klasifikasi sentimen terkait kasus korupsi PT Pertamina. Pada penelitian ini, proses pelatihan diatur dengan batch size = 16 dan epochs = 3, sehingga seluruh data training diproses sebanyak tiga kali. Optimizer yang digunakan adalah AdamW dengan learning rate sebesar 2e-5[10], karena algoritma ini

dirancang untuk menjaga stabilitas pembaruan bobot pada arsitektur Transformer seperti BERT.

Pengaturan pelatihan didefinisikan melalui TrainingArguments, yang mencakup strategi evaluasi setiap epoch, penyimpanan model terbaik berdasarkan metrik F1-score, penggunaan weight decay untuk regularisasi, logging setiap 50 langkah, serta opsi memuat model terbaik di akhir proses. Strategi ini memungkinkan model untuk belajar lebih efektif, mencegah overfitting, serta memastikan model terbaik dapat digunakan pada tahap pengujian.

G. Tokenisasi IndoBert

Sebelum masuk ke tahap pelatihan, data teks terlebih dahulu melalui proses tokenisasi menggunakan tokenizer IndoBERT. Tokenisasi ini berfungsi untuk mengubah teks menjadi representasi numerik yang dapat dipahami oleh model. Hasil dari tokenisasi mencakup *input_ids* yang merepresentasikan indeks token hasil pemetaan kata ke dalam vocabulary IndoBERT. Sebagai contoh, kalimat “*PT Pertamina (Persero) terus memperkuat pondasi bisnisnya sebagai langkah strategis untuk memastikan pertumbuhan yang berkelanjutan di tengah dinamika industri energi global*” setelah melalui proses tokenisasi menghasilkan keluaran berupa token: `['[CLS]', 'pt', 'pertamina', 'persero', 'terus', 'memperkuat', 'pondasi', 'bisnisnya', 'sebagai', 'langkah', 'strategis', 'untuk', 'memastikan', 'pertumbuhan', 'yang', 'berkelanjutan', 'di', 'tengah', 'dinamika', 'industri', 'energi', 'global', '[SEP]']`. Token tersebut kemudian dikonversi menjadi *input_ids*: [2, 914, 6414, 4413, 944, 6789, 12249, 13731, 242, 1775, 4581, 90, 4052, 2471, 34, 8693, 26, 1172, 11035, 1934, 2479, 4092, 3275, 3]. Selanjutnya, dataset hasil tokenisasi ini dikonversi ke dalam format PyTorch tensor menggunakan perintah `with_format("torch")` agar kompatibel dengan modul Trainer pada library HuggingFace. Dataset yang telah diproses kemudian dibagi menjadi tiga bagian, yaitu *train_dataset* untuk tahap pelatihan, *eval_dataset* untuk tahap validasi, dan *test_dataset* untuk tahap pengujian. Pembagian ini memastikan proses pelatihan dapat berlangsung secara terstruktur, dengan adanya data validasi untuk memantau kinerja selama pelatihan, serta data uji untuk mengevaluasi performa model secara menyeluruh setelah proses fine-tuning selesai dilakukan.

H. Training dan Validasi Data

Proses training dilakukan dengan memanfaatkan modul Trainer dari pustaka Hugging Face, yang secara otomatis menangani perulangan epoch, pembaruan bobot, serta evaluasi model. Pada tahap training, model menerima data dalam bentuk batch dan menghitung loss berdasarkan selisih antara prediksi dengan label sebenarnya. Optimizer yang digunakan adalah AdamW, dengan proses pembaruan bobot berlangsung selama tiga epoch untuk memastikan model dapat mempelajari pola representasi sentimen dari dataset secara optimal.

Selama proses fine-tuning, sistem menampilkan log teknis seperti pemuatan file `pytorch_model.bin` sebesar 498 MB,

peringatan terkait penggunaan versi lama AdamW, serta inisialisasi ulang pada layer `classifier.bias` dan `classifier.weight` yang memang perlu dilatih dari awal. Selain itu, terdapat juga pesan peringatan lain seperti penggunaan `evaluation_strategy` yang sudah usang serta konfigurasi `pin_memory`, namun hal ini tidak memengaruhi jalannya training.

Pada epoch pertama, model memperoleh training loss sebesar 0,4498 dan validation loss sebesar 0,3224, dengan tingkat akurasi mencapai 85,57%. Hasil ini menunjukkan bahwa meskipun model masih dalam tahap awal pembelajaran, ia sudah mampu mengenali pola sentimen dengan cukup baik.

Memasuki epoch kedua, kinerja model semakin meningkat dengan training loss turun menjadi 0,2778 dan validation loss menurun ke 0,2934. Akurasi pada tahap ini naik menjadi 87,06%, dengan nilai precision, recall, dan F1-score yang relatif seimbang, menandakan kemampuan generalisasi model semakin baik.

Pada epoch ketiga, training loss kembali turun signifikan menjadi 0,1353, sementara validation loss sedikit meningkat ke angka 0,3498. Walaupun begitu, akurasi justru meningkat hingga mencapai 89,05% dengan nilai precision dan F1-score di kisaran 0,891, menunjukkan bahwa model mampu mencapai performa yang stabil meskipun terdapat sedikit fluktuasi pada validation loss.

Secara keseluruhan, hasil training memperlihatkan tren positif di mana akurasi model terus meningkat dari epoch pertama hingga ketiga. Performa akhir dengan akurasi 89,05% dan F1-score 0,8911 menunjukkan bahwa model IndoBERT yang telah melalui proses fine-tuning berhasil belajar secara efektif untuk mengklasifikasikan sentimen publik terhadap isu korupsi pada BUMN strategis, sehingga diharapkan dapat memberikan prediksi yang lebih akurat dalam analisis opini publik.

I. Non Fine-Tuning (Ekstraksi Embedding)

Pada pendekatan non fine-tuning, IndoBERT tidak mengalami pembaruan bobot selama proses pelatihan, melainkan hanya berfungsi sebagai feature extractor untuk menghasilkan representasi numerik dari teks atau yang disebut embedding. Proses ini dimulai dari data teks yang telah ditokenisasi menggunakan IndoBERT tokenizer, sehingga diperoleh tiga komponen utama yaitu *input_ids*, *attention_mask*, dan *token_type_ids*. Ketiga komponen ini kemudian dimasukkan ke dalam model IndoBERT untuk menghasilkan *hidden states*, yakni representasi vektor dari setiap token dalam kalimat. Dari *hidden states* tersebut, embedding dapat diekstraksi dengan beberapa cara, seperti mengambil representasi token [CLS] pada lapisan terakhir atau melakukan rata-rata (*mean pooling*) pada seluruh token [12], [13].

Embedding yang diperoleh tidak digunakan langsung untuk klasifikasi oleh IndoBERT, melainkan diekspor sebagai fitur ke dalam algoritma pembelajaran mesin tradisional, seperti Logistic Regression dan Linear SVM. Dengan cara ini,

IndoBERT berperan sebagai penyedia representasi bahasa yang kaya konteks, sementara algoritma klasik bertugas melakukan klasifikasi berdasarkan fitur tersebut.

Pendekatan ini memberikan efisiensi komputasi karena tidak memerlukan proses pembaruan bobot yang kompleks. Selain itu, metode non fine-tuning juga memisahkan tahap ekstraksi fitur dan klasifikasi secara jelas, serta mempermudah perbandingan performa dengan model IndoBERT yang di-fine-tune. Walaupun performa model non fine-tuning biasanya sedikit lebih rendah dibandingkan model yang di-fine-tune, metode ini tetap relevan karena lebih ringan secara sumber daya dan cepat diimplementasikan.

J. Classifier

Tahap klasifikasi merupakan inti dari penelitian ini, di mana embedding yang dihasilkan oleh IndoBERT digunakan sebagai representasi fitur untuk memprediksi sentimen publik. Dua algoritma yang dipilih adalah Logistic Regression dan Linear Support Vector Machine (SVM), keduanya dikenal efektif dalam menangani data teks berdimensi tinggi seperti embedding dari model bahasa.

1) Logistic Regression

Logistic Regression termasuk algoritma klasifikasi berbasis probabilistik. Alih-alih memprediksi nilai kontinu seperti pada regresi linear, model ini menghitung probabilitas suatu data masuk ke kelas tertentu. Fungsi dasarnya menggunakan kombinasi linear dari vektor fitur x dengan bobot w dan bias b , kemudian diproyeksikan melalui fungsi sigmoid [6], [10]. Rumus probabilitas kelas positif dituliskan sebagai:

$$P(y=1|x) = \sigma(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}}$$

Dalam kasus klasifikasi sentimen biner (positif dan negatif), nilai probabilitas mendekati 1 menunjukkan kecenderungan sentimen positif, sedangkan mendekati 0 menunjukkan sentimen negatif. Untuk melatih model, Logistic Regression meminimalkan fungsi *binary cross-entropy loss* yang dirumuskan sebagai:

$$L = -\frac{1}{N} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

dengan y_i adalah label sebenarnya, \hat{y}_i adalah probabilitas hasil prediksi, dan N adalah jumlah data [11], [12]. Fungsi ini memastikan agar hasil prediksi semakin dekat dengan label asli. Kelebihan Logistic Regression adalah interpretabilitasnya yang baik karena model menghasilkan nilai probabilitas, sehingga cocok digunakan dalam analisis opini publik.

2) Linear SVM

Berbeda dengan Logistic Regression, Linear SVM berfokus pada pencarian hyperplane terbaik yang dapat memisahkan data ke dalam dua kelas dengan margin maksimum. Fungsi diskriminan Linear SVM dituliskan sebagai [9], [12]:

$$f(x) = w^T x + b$$

Keputusan klasifikasi dilakukan berdasarkan tanda dari $f(x)$: jika $f(x) > 0$ maka data dikategorikan sebagai kelas positif, sedangkan jika $f(x) < 0$ maka masuk ke kelas negatif. Selama pelatihan, Linear SVM menggunakan hinge loss yang didefinisikan sebagai:

$$L = \frac{1}{N} \sum_{i=1}^N \max(0, 1 - y_i (w^T x_i + b)) + \lambda \|w\|^2$$

dengan $y_i \in \{-1, +1\}$ sebagai label kelas, λ parameter regulasi, dan $\|w\|^2$ berfungsi untuk mengendalikan margin agar tetap besar serta mencegah overfitting [6], [12], [13]. Linear SVM sangat efektif dalam menangani data berukuran besar dengan banyak dimensi, seperti embedding IndoBERT [10].

K. Evaluasi Model

Tahap evaluasi model dilakukan setelah proses training dan validasi selesai, dengan tujuan untuk mengukur sejauh mana model mampu mengklasifikasikan sentimen publik secara akurat. Evaluasi menggunakan test dataset, yaitu data yang tidak pernah digunakan selama pelatihan, sehingga hasilnya dapat mencerminkan performa model secara objektif [1], [3], [8]. Dalam penelitian ini digunakan dua pendekatan, yaitu IndoBERT dengan fine-tuning dan IndoBERT tanpa fine-tuning (menggunakan Logistic Regression atau Linear SVM sebagai classifier) [6], [10], [12].

Pada IndoBERT dengan fine-tuning, bobot model dilatih ulang menggunakan dataset sentimen sehingga representasi bahasa menjadi lebih sesuai dengan konteks opini publik terkait PT Pertamina. Sementara itu, pada IndoBERT tanpa fine-tuning, model hanya digunakan sebagai feature extractor, yaitu menghasilkan representasi vektor teks yang kemudian diklasifikasikan menggunakan algoritma machine learning klasik [9], [11]. Kedua pendekatan ini kemudian dievaluasi dengan metrik yang sama untuk memperoleh perbandingan performa secara adil [18], [19].

Evaluasi model berbasis klasifikasi umumnya digambarkan menggunakan confusion matrix, yaitu tabel dua dimensi yang menunjukkan perbandingan antara prediksi model dengan label aktual [7], [14], [17]. Komponennya adalah sebagai berikut:

- True Positive (TP): Jumlah data sentimen positif yang diprediksi benar sebagai positif.
- Negative (TN): Jumlah data sentimen negatif yang diprediksi benar sebagai negatif.
- False Positive (FP): Jumlah data sentimen negatif yang salah diprediksi sebagai positif.

- False Negative (FN): Jumlah data sentimen positif yang salah diprediksi sebagai *negative*[9].

TABEL I
STRUKTUR UMUM CONFUSION MATRIX

Kelas aktual	Prediksi positif	Prediksi negatif
Positif	TP	FN
Negatif	FN	TN

Dari table ini, semua data metrix dapat di hitung dengan rumus berikut :

1) *Akurasi (Accuracy)*

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Akurasi menghitung proporsi prediksi yang benar (baik positif maupun negatif) dibandingkan dengan total seluruh data uji. Nilai akurasi tinggi menunjukkan model mampu membuat prediksi yang tepat secara keseluruhan. Namun, akurasi saja tidak cukup bila data tidak seimbang, karena model bisa hanya fokus pada kelas mayoritas [5], [18].

2) *Presisi (Precision)*

$$\text{Precision} = \frac{TP}{TP + FP}$$

Presisi menunjukkan seberapa tepat model dalam memberikan prediksi positif. Nilai tinggi berarti sebagian besar data yang diprediksi positif memang benar-benar positif, sehingga kesalahan *false positive* sangat kecil. Presisi penting pada kasus di mana kesalahan positif palsu perlu diminimalkan [6], [10].

3) *Recall (Sensitivity)*

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall mengukur sejauh mana model mampu menemukan semua data positif. Nilai recall tinggi berarti sebagian besar data positif berhasil dikenali model, dan kesalahan *false negative* kecil. Recall penting pada kasus di mana kehilangan data positif (*missed detection*) berisiko besar[11], [12].

4) *F1-Score*

$$F1 = 2x \frac{Precision \times Recall}{Precision + Recall}$$

F1-score adalah rata-rata harmonik dari presisi dan recall. Metrik ini digunakan untuk mencari keseimbangan antara keduanya. Nilai F1-score tinggi menunjukkan bahwa model tidak hanya tepat (precision tinggi) tetapi juga menyeluruh dalam deteksi (recall tinggi). Metrik ini sangat relevan pada dataset yang memiliki distribusi kelas tidak seimbang [9], [13].

III. HASIL DAN PEMBAHASAN

Pada pembahasan hasil penelitian ini, dataset diambil dari media sosial X (dulu Twitter) melalui skrip Python yang dikembangkan secara mandiri, sebagaimana dapat dilihat pada Tabel 2. Total dataset yang diperoleh adalah 2012 tweet terkait isu korupsi pada BUMN strategis. Data yang sudah dikumpulkan kemudian diolah terlebih dahulu melalui tahap preprocessing, meliputi pembersihan teks, normalisasi, tokenisasi, dan stemming, agar hasil modeling menjadi lebih optimal.

Bab ini menyajikan hasil implementasi analisis sentimen dengan menggunakan algoritma IndoBERT, baik melalui fine-tuning maupun tanpa fine-tuning, serta dibandingkan dengan algoritma klasik seperti Logistic Regression dan Linear SVM. Pembahasan difokuskan pada dua aspek, yaitu evaluasi kuantitatif melalui metrik performa model (akurasi, precision, recall, F1-score, serta confusion matrix), dan evaluasi kualitatif melalui analisis distribusi sentimen dan pola opini publik yang muncul dari dataset.

Hasil penelitian dijelaskan secara terstruktur, dimulai dari deskripsi dan eksplorasi dataset, proses preprocessing, pelatihan dan evaluasi model, hingga analisis perbandingan performa antar model. Dengan struktur ini, Bab ini tidak hanya menekankan performa numerik antar model, tetapi juga memberikan wawasan mengenai persepsi publik terhadap isu korupsi pada BUMN strategis, serta memperkuat relevansi pendekatan berbasis IndoBERT dalam analisis sentimen berbahasa Indonesia, khususnya pada dataset yang memiliki ketidakseimbangan kelas.

A. Deskripsi Dataset

Dataset penelitian ini diperoleh dari media sosial X (dulu Twitter) melalui skrip Python yang dikembangkan secara mandiri. Proses pengumpulan data difokuskan pada tweet yang membahas isu korupsi pada BUMN strategis pada tahun 2025, sehingga diperoleh total 2.012 tweet yang relevan. Setiap tweet kemudian dikategorikan menjadi dua kelas sentimen, yaitu positif dan negatif, untuk mempermudah analisis opini publik terhadap perusahaan.

TABEL 2
DISTRIBUSI DATASET SEBELUM SMOTE

Sentimen	Jumlah tweet	Presentase (%)
Negatif	1.408	69,93 %
Positif	604	30,07 %
Total	2.012	100 %

Dari tabel 2 tersebut, terlihat bahwa sentimen negatif mendominasi dengan 1.408 tweet, sedangkan sentimen positif hanya terdiri dari 604 tweet. Kondisi ini menunjukkan adanya ketidakseimbangan kelas, yang dapat memengaruhi performa model karena model cenderung bias terhadap kelas mayoritas dan kurang mampu mengenali pola pada kelas minoritas.

Selain distribusi kelas, dataset memiliki karakteristik lain yang perlu diperhatikan. Panjang teks tweet bervariasi, mulai dari beberapa kata hingga beberapa kalimat. Bahasa yang

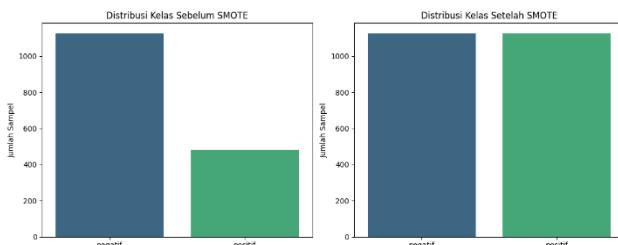
digunakan bersifat informal, sering mengandung slang, singkatan, simbol, hashtag, atau mention pengguna lain. Karakteristik ini memerlukan tahap preprocessing yang teliti, termasuk pembersihan teks, normalisasi, tokenisasi, dan stemming, agar model dapat memproses data secara optimal.

Untuk mengatasi ketidakseimbangan kelas, penelitian ini menerapkan SMOTE (*Synthetic Minority Oversampling Technique*). SMOTE menambahkan sampel sintetis pada kelas minoritas (positif) sehingga jumlah tweet di kelas positif dan negatif menjadi seimbang, masing-masing 1.126 tweet, seperti ditampilkan pada tabel 3.

TABEL 3
DISTRIBUSI DATASET SE SMOTE

Sentimen	Jumlah tweet	Persentase (%)
Negatif	1.126	50 %
Positif	1.126	50 %
Total	2.252	100 %

Dengan penerapan SMOTE, model analisis sentimen dapat belajar secara optimal pada kedua kelas, meningkatkan kemampuan prediksi pada kelas minoritas, dan menghasilkan performa evaluasi yang lebih akurat. Dengan pemahaman mendalam tentang karakteristik dataset ini dan penerapan SMOTE, tahap preprocessing dan modeling dapat dilakukan secara lebih efektif, serta analisis sentimen dapat mencerminkan persepsi publik terhadap isu korupsi secara lebih adil dan representatif.



Gambar 3. Distribusi dataset sebelum dan setelah SMOTE

B. Analisis Hasil Klasifikasi

Analisis hasil klasifikasi dilakukan untuk mengukur performa masing-masing model dalam menganalisis sentimen publik terhadap isu korupsi. Tiga pendekatan yang digunakan adalah IndoBERT dengan fine-tuning penuh, IndoBERT (Non Fine-Tuning) dengan Logistic Regression, dan IndoBERT (Non Fine-Tuning) dengan Linear SVM.

Hasil pengujian menunjukkan bahwa model IndoBERT dengan fine-tuning memperoleh performa terbaik dengan nilai akurasi sebesar 82,67% serta F1-Score sebesar 82,35%. Hal ini membuktikan bahwa fine-tuning mampu menyesuaikan parameter model dengan data penelitian, sehingga menghasilkan representasi yang lebih kontekstual dan relevan.

Sementara itu, model IndoBERT (Non Fine-Tuning) + Logistic Regression hanya mencapai akurasi sebesar 80,20%

dengan F1-Score 79,52%. Meskipun performanya cukup baik, nilai presisi dan recall pada kelas positif masih lebih rendah dibandingkan fine-tuning, sehingga model ini cenderung kurang optimal dalam mengklasifikasikan sentimen positif.

Adapun model IndoBERT (Non Fine-Tuning) + Linear SVM menghasilkan akurasi terendah yaitu 78,22% dengan F1-Score 77,62%. Hal ini menunjukkan bahwa pendekatan tanpa fine-tuning membuat representasi fitur yang dihasilkan IndoBERT belum sepenuhnya termanfaatkan oleh algoritma klasifikasi lain, sehingga performanya menurun dibandingkan dengan dua model sebelumnya.

Secara keseluruhan, hasil perbandingan dapat dilihat pada Tabel 4 berikut:

TABEL 4
HASIL PERBANDINGAN KINERJA MODEL INDOBERT FINE-TUNING DAN INDOBERT NON-FT

	Model		
	IndoBert + Fine-Tuning	IndoBert (non-ft) + Logistic Regression	IndoBert (Non- FT) + Linear SVM
Akurasi	82,67 %	80,20 %	78,22 %
Presisi	82,27 %	79,51 %	77,45 %
Recall	82,675 %	80,20 %	78,225 %
F1-Score	82,35 %	79,52 %	77,62 %
Presisi Negatif	85,81 %	83,01 %	82,12 %
Recall Negatif	90,07 %	90,07 %	87,94 %
Presisi Positif	74,07 %	71,43 %	66,67 %
Recall Positif	65,57 %	57,38 %	55,74 %

Berdasarkan tabel 4 di atas, IndoBERT dengan Fine-Tuning menunjukkan performa paling konsisten, dengan sebagian besar nilai berada di kisaran 82–85%, menunjukkan kestabilan dan adaptasi yang baik terhadap konteks bahasa Indonesia. Model Logistic Regression cukup baik dalam mengenali kelas negatif, tetapi performa pada kelas positif lebih rendah. Linear SVM memiliki performa paling rendah dan cenderung fluktuatif, terutama pada klasifikasi kelas positif.

Secara keseluruhan, tren ini menegaskan bahwa fine-tuning memberikan keuntungan signifikan, baik dari sisi akurasi maupun kestabilan prediksi, terutama dalam menyeimbangkan klasifikasi antara kelas negatif dan positif.

C. Evaluasi Confusion Matrix

Analisis confusion matrix dilakukan untuk menggali lebih dalam kinerja model dalam mengklasifikasikan sentimen publik terhadap isu korupsi. Confusion matrix memberikan gambaran mengenai jumlah prediksi yang benar (True Positive/True Negative) serta jumlah prediksi yang salah (False Positive/False Negative) pada setiap kelas. Dengan

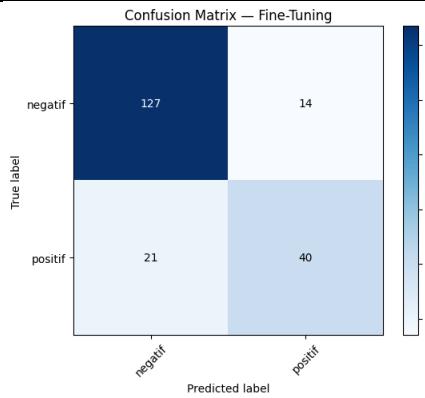
menggunakan confusion matrix, kelemahan dan kekuatan model dapat terlihat lebih jelas dibandingkan hanya mengandalkan nilai akurasi semata.

1) Confusion Matrix IndoBert Fine-Tuning

Model IndoBERT dengan fine-tuning menunjukkan performa terbaik dengan akurasi 82,67%. Dari 141 data negatif, 127 berhasil diprediksi dengan benar, sementara 14 salah diklasifikasikan sebagai positif. Pada kelas positif, 40 dari 61 data terprediksi dengan benar, meskipun terdapat beberapa false negative. Hal ini menunjukkan model mampu menangkap pola sentimen secara kontekstual dan menyeimbangkan prediksi antara kelas negatif dan positif.

TABEL 5
CONFUSION MATRIX INDOBERT FINE-TUNING

	Pred : Negatif	Pred : Positif	Jumlah
Asli Negatif	127	14	141
Asli Positif	21	40	61



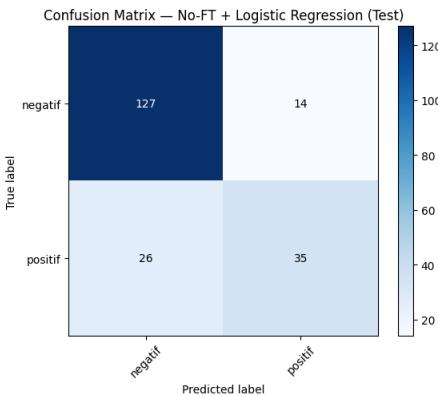
Gambar 4. Confusion matrix model IndoBERT fine-tuning

2) Confusion Matrix IndoBert Non Fine-Tuning (Logistic Regression)

Model ini memiliki akurasi 80,20% dan tetap cukup baik dalam mengenali kelas negatif (127 prediksi benar dari 141 data). Namun, pada kelas positif, jumlah false negative meningkat (26 data positif salah diprediksi sebagai negatif), menunjukkan kemampuan model dalam mengklasifikasikan sentimen positif lebih rendah dibanding Fine-Tuning.

TABEL 6
CONFUSION MATRIX INDOBERT NON-FT + LOGISTICS REGRESSION

	Pred : Negatif	Pred : Positif	Jumlah
Asli Negatif	127	14	141
Asli Positif	26	35	61



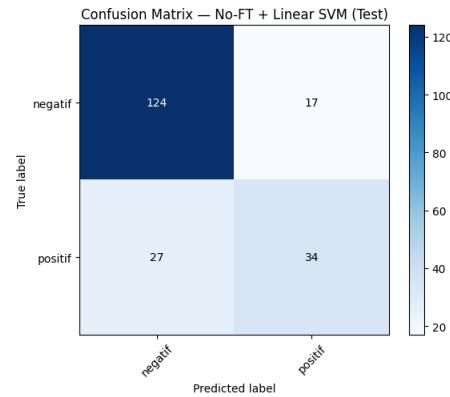
Gambar 5. Confusion matrix model IndoBERT non-FT + Logistic Regression

3) Confusion Matrix IndoBert Non Fine-Tuning (Linear SVM)

Model ini menunjukkan akurasi terendah, yaitu 78,21%. Meskipun prediksi pada kelas negatif masih cukup baik (127 data benar dari 141), kesalahan pada kelas positif semakin meningkat (26 data positif salah diklasifikasikan), menandakan Linear SVM kurang efektif menangkap representasi semantik dari IndoBERT tanpa fine-tuning.

TABEL 7
CONFUSION MATRIX INDOBERT NON-FT + LINEAR SVM

	Prediksi Negatif	Prediksi Positif	Jumlah
Asli Negatif	127	14	141
Asli Positif	26	35	61



Gambar 6. Confusion matrix model IndoBERT non-FT + Linear SVM

4) Ringkasan Perbandingan

Secara umum, hasil penelitian menunjukkan bahwa model IndoBERT dengan fine-tuning memberikan performa terbaik dibandingkan dua model non-fine-tuning. Fine-tuning membuat model lebih adaptif

terhadap konteks bahasa Indonesia dalam dataset 2.012 tweet, sehingga mampu menyeimbangkan prediksi antara kelas negatif dan positif. Dari analisis confusion matrix (Table 5), model ini berhasil mengklasifikasikan 127 data negatif dan 40 data positif dengan benar, sementara terdapat 14 false positive pada kelas negatif dan 21 false negative pada kelas positif. Meskipun masih ada beberapa kesalahan, model ini secara keseluruhan mampu menangkap opini publik dengan representasi yang lebih akurat.

Model IndoBERT non-fine-tuning + Logistic Regression masih cukup baik dalam mengklasifikasikan kelas negatif (127 data negatif benar), tetapi memiliki kelemahan pada kelas positif dengan 26 false negative (35 positif benar), menunjukkan keterbatasan model dalam mengenali opini minoritas. Sementara itu, model IndoBERT non-fine-tuning + Linear SVM memiliki performa terendah, dengan jumlah positif benar 35 dan false negative 26, menegaskan bahwa representasi fitur dari model tanpa fine-tuning belum dimanfaatkan secara optimal oleh algoritma klasifikasi ini.

TABEL 8
RINGKASAN PEANDINGAN CONFUSION MATRIX INDOBERT

Model	Akurasi	Kekuatan	Kelemahan
IndoBert Fine Tuning	82,67 %	Tingkat prediksi benar tinggi pada kelas negatif & positif	Masih ada FN pada kelas positif
IndoBert Non-FT+Logistic Regression	80,20 %	Mampu mengenali sebagian besar kelas negatif	FN pada kelas positif relatif lebih tinggi
IndoBert Non-FT+Linear SVM	78,21 %	Cukup baik di kelas negatif	Kesalahan lebih besar pada kelas positif

Setelah melihat Table 8, dapat disimpulkan bahwa fine-tuning pada IndoBERT secara signifikan meningkatkan kemampuan model dalam menyeimbangkan prediksi sentimen positif dan negatif, yang sejalan dengan temuan ROC Curve dan analisis wordcloud. Precision tinggi pada model fine-tuning menunjukkan bahwa prediksi positif sebagian besar benar, sedangkan recall tinggi menandakan model mampu menangkap hampir seluruh data positif. Kombinasi precision dan recall yang baik tercermin pada F1-score, yang menegaskan keunggulan model dalam analisis sentimen pada dataset polar. Temuan ini menegaskan bahwa fine-tuning memberikan keuntungan signifikan dibandingkan metode

non-fine-tuning, baik dari sisi akurasi keseluruhan maupun kemampuan menangani ketidakseimbangan kelas.

D. Evaluasi Hasil Model (Visualisasi)

Tahap evaluasi hasil model tidak hanya dilakukan melalui perhitungan metrik kuantitatif, tetapi juga menggunakan pendekatan visualisasi untuk memberikan pemahaman yang lebih intuitif mengenai pola data dan performa klasifikasi. Dua jenis visualisasi yang digunakan dalam penelitian ini adalah wordcloud sentimen dan ROC curve.

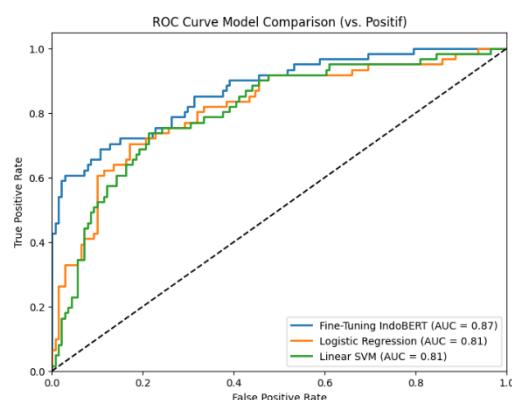


Gambar 7. Wordcloud sentimen positif dan negatif

Wordcloud pada sentimen positif memperlihatkan kata-kata dominan seperti “energi”, “terima kasih”, dan “kelola”. Kata-kata ini menunjukkan bahwa masyarakat memberikan apresiasi terhadap kinerja dan layanan BUMN strategis, terutama dalam hal pengelolaan energi, transparansi operasional, dan pelayanan publik. Kehadiran kata “terima kasih” menandakan adanya pengakuan dan kepuasan pengguna terhadap pelayanan atau program yang dijalankan oleh perusahaan. Secara keseluruhan, sentimen positif mencerminkan persepsi masyarakat yang mendukung atau menghargai kinerja BUMN strategis, menunjukkan adanya elemen kepercayaan dan penghargaan terhadap tata kelola yang dianggap baik.

Sebaliknya, pada wordcloud sentimen negatif tampak dominasi kata “korupsi”, “minyak mentah”, “kejaksaan agung”, dan “tata kelola”. Kata-kata ini mengindikasikan adanya opini publik yang kritis terhadap isu dugaan korupsi dan masalah pengelolaan perusahaan. Misalnya, kata “korupsi” dan “kejaksaan agung” menunjukkan perhatian masyarakat terhadap tindakan hukum dan investigasi terkait dugaan penyalahgunaan wewenang, sedangkan kata “tata kelola” menandakan kritik terhadap aspek manajemen dan transparansi operasional. Sentimen negatif ini mencerminkan adanya kekhawatiran, ketidakpuasan, atau skeptisme publik terhadap kinerja BUMN strategis, sehingga memberikan insight penting terkait isu tata kelola dan integritas.

Adapun untuk wordcloud sentimen netral tidak tersedia karena dataset tidak memiliki cukup teks dalam kategori tersebut. Dengan demikian, dapat disimpulkan bahwa percakapan publik mengenai isu korupsi pada BUMN strategis di media sosial lebih banyak bersifat polar, dengan dominasi opini yang cenderung positif atau negatif, sementara opini netral relatif minim. Analisis ini menunjukkan bahwa masyarakat memiliki persepsi yang kuat terhadap isu korupsi dan tata kelola BUMN strategis, baik dalam bentuk dukungan maupun kritik.



Gambar 8. Hasil ROC curve model IndoBERT

Kurva ROC memperlihatkan bahwa model Fine-Tuning IndoBERT memiliki nilai Area Under Curve (AUC) sebesar 0,87, yang menunjukkan kemampuan model yang sangat baik dalam membedakan kelas positif dan negatif. Sementara itu, model Logistic Regression dan Linear SVM masing-masing memperoleh nilai AUC sebesar 0,81, yang menunjukkan performa yang cukup baik tetapi masih di bawah kinerja Fine-Tuning IndoBERT. Nilai AUC ini membantu pembaca memahami sejauh mana model dapat memisahkan kelas sentimen: nilai 0,5 menandakan model acak, sedangkan 1,0 menunjukkan pemisahan sempurna. Dengan perbedaan 0,06 antara Fine-Tuning IndoBERT dan model klasik, dapat disimpulkan bahwa proses fine-tuning memberikan peningkatan performa yang signifikan, memungkinkan model memahami konteks teks secara lebih mendalam dibandingkan metode tradisional.

Secara keseluruhan, visualisasi wordcloud dan ROC curve memperkuat hasil evaluasi kuantitatif. Wordcloud berhasil menunjukkan kata-kata dominan pada masing-masing kategori sentimen, sedangkan ROC curve menegaskan bahwa Fine-Tuning IndoBERT merupakan model terbaik dalam penelitian ini. Dengan demikian, kombinasi kedua visualisasi ini memberikan gambaran komprehensif mengenai performa model dalam menganalisis sentimen publik terhadap isu korupsi pada BUMN strategis.

E. Analisis Performa Model dan Batasan Data

Hasil penelitian menunjukkan bahwa model IndoBERT dengan fine-tuning memberikan performa terbaik dibandingkan model tradisional seperti Logistic Regression dan Linear SVM. Hal ini terlihat dari perolehan akurasi dan F1-score yang lebih tinggi, serta nilai AUC sebesar 0,87 yang menunjukkan kemampuan diskriminatif model dalam membedakan kelas positif dan negatif secara lebih efektif. Nilai AUC ini membantu pembaca memahami sejauh mana model dapat memisahkan kelas sentimen: nilai 0,5 menandakan model acak, sedangkan nilai 1,0 menunjukkan pemisahan sempurna. Dengan selisih 0,06 dibandingkan model Logistic Regression dan Linear SVM (AUC 0,81), dapat disimpulkan bahwa fine-tuning memberikan

peningkatan performa signifikan, memungkinkan model memahami konteks teks secara lebih mendalam. Visualisasi ROC Curve juga memperkuat temuan ini, di mana kurva IndoBERT lebih menjauhi garis diagonal dibandingkan dua model lainnya, menegaskan keunggulan performa diskriminatif model berbasis Transformer.

Hasil wordcloud memberikan wawasan lebih lanjut mengenai pola opini publik pada 2012 tweet terkait isu korupsi pada BUMN strategis. Wordcloud sentimen positif menunjukkan kata-kata dominan seperti “energi”, “terima kasih”, dan “kelola”, yang mencerminkan apresiasi masyarakat terhadap kinerja dan pengelolaan BUMN strategis, terutama dalam layanan publik, pengelolaan energi, dan transparansi operasional. Sebaliknya, wordcloud sentimen negatif didominasi kata-kata seperti “korupsi”, “minyak mentah”, “kejaksaan agung”, dan “tata kelola”, yang menunjukkan kritik publik terhadap dugaan korupsi, manajemen, dan tata kelola perusahaan. Karena sentimen netral tidak dianalisis, visualisasi wordcloud hanya fokus pada polaritas positif dan negatif, sehingga memberikan gambaran opini publik yang dominan namun lebih tajam terhadap sentimen yang menonjol.

Terdapat beberapa keterbatasan yang perlu dicatat. Pertama, distribusi data sentimen tidak seimbang, dengan jumlah tweet positif dan negatif yang berbeda, sehingga analisis pola opini publik menjadi lebih terfokus pada kelompok mayoritas. Kedua, jumlah sampel 2012 tweet, meskipun relatif besar dibanding penelitian serupa, mungkin belum sepenuhnya mewakili opini masyarakat secara luas, sehingga generalisasi model perlu hati-hati jika diterapkan pada konteks BUMN strategis lain. Ketiga, meskipun IndoBERT menunjukkan performa terbaik, model ini membutuhkan sumber daya komputasi lebih tinggi dibanding Logistic Regression dan Linear SVM, sehingga pertimbangan penting jika akan diimplementasikan pada lingkungan produksi dengan keterbatasan infrastruktur.

Dengan demikian, meskipun penelitian menegaskan keunggulan IndoBERT dalam analisis sentimen, perlu diperhatikan keterbatasan terkait ketidakseimbangan data, jumlah sampel, serta kebutuhan komputasi yang tinggi. Penelitian selanjutnya dapat memperluas jumlah dataset, menyeimbangkan distribusi kelas sentimen, serta mengeksplorasi teknik optimasi model agar lebih efisien. Fokus pada analisis sentimen positif dan negatif tetap relevan, mengingat sentimen netral tidak dianalisis, sehingga memberikan gambaran opini publik yang jelas mengenai isu korupsi pada BUMN strategis.

F. Analisis Perbandingan dengan Penelitian Sebelumnya

Analisis perbandingan dilakukan untuk menempatkan hasil penelitian ini dalam konteks literatur yang sudah ada. Dengan cara ini, dapat terlihat posisi kontribusi penelitian ini apakah sejalan, lebih baik, atau lebih rendah dibandingkan penelitian sebelumnya. Perbandingan dilakukan terutama dari aspek akurasi, F1-score, recall, serta AUC yang relevan dengan kajian sentimen publik.

TABEL 9
PERBANDINGAN DENGAN PENELITIAN SEBELUMNYA

Penelitian n	Metode	Hasil penelitian sebelumnya	Hasil penelitian saat ini	Perbandingan
Irmawan dkk.[3]	Logistic Regression + TF-IDF	Akurasi 89%, F1 86,5%	Akurasi 82,67%, F1 82,35%	Lebih rendah; data formal lebih mudah dibanding opini media sosial.
Tarigan dkk.[6]	SVM	Akurasi 82%	Akurasi 82,67%	Konsisten; IndoBERT sedikit lebih unggul pada isu korupsi.
Jayadian ti dkk.[14]	IndoBert + R-CNN	Akurasi > 85%	Akurasi 82,67%, Recall Positif 65,57%	Lebih rendah; pola recall positif tetap lebih lemah dibanding negatif.
Al Farizi & Sibaroni [8]	BiLSTM + IndoBert Fine-Tuning	AUC > 0,85	AUC 0,87	Konsisten; fine-tuning meningkatkan kemampuan diskriminatif.
Kusoema & Ibrahim [2]	IndoBert + RCNN	Akurasi > 80%	Akurasi 82,67%	Sama-sama menunjukkan dominasi isu korupsi dan governance.

Berdasarkan tabel tersebut, terlihat bahwa akurasi penelitian ini lebih rendah dibandingkan Irmawan dkk.[3], namun sejalan dengan Tarigan dkk.[6]. Temuan Jayadianti dkk.[14] juga menguatkan hasil ini, di mana recall positif lebih lemah dibanding recall negatif, suatu pola yang juga muncul pada penelitian ini. Dari sisi AUC, penelitian ini memperoleh 0,87, konsisten dengan temuan Al Farizi & Sibaroni [8] yang melaporkan AUC di atas 0,85 dengan fine-tuning IndoBERT. Analisis wordcloud dalam penelitian ini juga mendukung temuan Kusoema & Ibrahim[2], yaitu dominasi isu governance dan korupsi dalam opini publik terkait Pertamina. Dengan demikian, penelitian ini menegaskan bahwa meskipun akurasi sedikit lebih rendah dibanding domain formal, hasilnya tetap kompetitif dan relevan untuk konteks media sosial.

IV. KESIMPULAN

Penelitian ini berhasil menganalisis sentimen publik terhadap isu korupsi pada BUMN strategis tahun 2025 dengan memanfaatkan 2.012 tweet dari media sosial X (Twitter). Tiga pendekatan digunakan dalam penelitian ini, yaitu IndoBERT dengan fine-tuning penuh, IndoBERT tanpa fine-tuning dengan Logistic Regression, dan IndoBERT tanpa fine-tuning dengan Linear SVM. Hasil penelitian menunjukkan bahwa model IndoBERT dengan fine-tuning memberikan performa terbaik, dengan akurasi sebesar 82,67%, F1-score 82,35%, dan nilai AUC 0,87. Nilai AUC ini menunjukkan kemampuan model dalam membedakan sentimen positif dan negatif secara sangat baik, serta memahami konteks bahasa Indonesia dengan lebih mendalam dibandingkan model tradisional. Sebaliknya, pendekatan tanpa fine-tuning hanya mencapai akurasi 80,20% dan 78,22%, dengan performa yang lebih rendah khususnya dalam mengenali sentimen positif. Temuan ini menegaskan bahwa proses fine-tuning pada model berbasis Transformer secara signifikan meningkatkan kemampuan analisis sentimen.

Analisis wordcloud memperlihatkan bahwa opini publik terhadap isu korupsi pada BUMN strategis bersifat polar. Sentimen positif didominasi kata-kata yang terkait dengan apresiasi terhadap pengelolaan energi, pelayanan publik, dan transparansi operasional, sedangkan sentimen negatif menonjolkan kata-kata terkait dugaan korupsi, manajemen, dan tata kelola perusahaan. Temuan ini sejalan dengan hasil ROC Curve yang menegaskan keunggulan IndoBERT fine-tuning dibandingkan model tradisional dalam membedakan kelas sentimen dan menghasilkan representasi fitur yang lebih kontekstual.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan, yaitu distribusi data sentimen yang tidak seimbang, jumlah sampel yang relatif terbatas, serta kebutuhan komputasi yang lebih tinggi untuk model IndoBERT. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas jumlah dataset, menyeimbangkan distribusi kelas sentimen, serta mengeksplorasi teknik optimasi model agar lebih efisien dan dapat diterapkan pada skala yang lebih besar. Penelitian berikutnya juga dapat mempertimbangkan penambahan analisis sentimen netral agar persepsi publik lebih komprehensif, memperluas cakupan dataset dari berbagai platform media sosial, dan menerapkan metode augmentasi data atau model Transformer terbaru untuk meningkatkan performa sambil mengurangi beban komputasi. Dengan demikian, penelitian ini menunjukkan bahwa penggunaan model berbasis Transformer dengan fine-tuning, seperti IndoBERT, efektif dalam menganalisis sentimen berbahasa Indonesia, khususnya untuk isu korupsi pada BUMN strategis, serta memberikan landasan bagi pengembangan sistem analisis sentimen yang lebih akurat dan representatif di masa depan.

DAFTAR PUSTAKA

- [1] R. A. Al Hazmi, "Pengaruh Korupsi Terhadap Pertumbuhan Ekonomi Indonesia," *Jurnal Acitya Ardana*, vol. 3, no. 2, pp. 85–92, Jun. 2024, doi: 10.31092/jaa.v3i2.2563.
- [2] W. J. Kusoma and I. Ibrahim, "Sentiment Analysis on the PT Pertamina Corruption Case using IndoBERT and RCNN Methods," *SISTEMASI*, vol. 14, no. 5, p. 2246, Sep. 2025, doi: 10.32520/stmsi.v14i5.5392.
- [3] O. A. Irmawan, I. Budi, A. B. Santoso, and P. K. Putra, "Improving Sentiment Analysis and Topic Extraction in Indonesian Travel App Reviews Through BERT Fine-Tuning," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 13, no. 2, pp. 359–370, Jul. 2024, doi: 10.23887/janapati.v13i2.77028.
- [4] N. A. Nevradia and M. A. Syaputra, "Sentiment Analysis of Telegram App Reviews on Google Play Store Using the Support Vector Machine (SVM) Algorithm," *Journal of Applied Informatics and Computing*, vol. 9, no. 1, pp. 96–105, Jan. 2025, doi: 10.30871/jaic.v9i1.8851.
- [5] A. R. Gunawan and R. F. Alfa Aziza, "Sentiment Analysis Using LSTM Algorithm Regarding Grab Application Services in Indonesia," *Journal of Applied Informatics and Computing*, vol. 9, no. 2, pp. 322–332, Mar. 2025, doi: 10.30871/jaic.v9i2.8696.
- [6] D. A. TARIGAN, Z. Situmorang, and R. Rosnely, "Analisis Sentimen Aplikasi Playstore Sirekap 2024 Pasca Pilpres Dengan Perbandingan Metode Support Vector Machine (SVM), Naïve Bayes Classifier Dan Random Forest," *Jurnal Teknologi Informatika dan Ilmu Komputer*, vol. 12, no. 3, pp. 661–670, Jun. 2025, doi: 10.25126/jtiik.2025129608.
- [7] R. Alif, A. Hazmi, P. Keuangan, and N. Stan, "Pengaruh Korupsi Terhadap Pertumbuhan Ekonomi Indonesia."
- [8] A. F. Al Farizi and Y. Sibaroni, "Implementation of BiLSTM and IndoBERT for Sentiment Analysis of TikTok Reviews," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 1, pp. 96–106, Jan. 2025, doi: 10.29100/jipi.v10i1.5815.
- [9] A. Yoga Pratama, G. Ananda Sanjaya, N. Khairunisa Lubis, and M. Rangga Aditya, "Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial," vol. 9, p. 2025, doi: 10.47002/metik.v9i1.1055.
- [10] U. Khairani, V. Mutiawani, and H. Ahmadian, "Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram," *Jurnal Teknologi Informatika dan Ilmu Komputer*, vol. 11, no. 4, pp. 887–894, Aug. 2024, doi: 10.25126/jtiik.1148315.
- [11] Erlin, Yulvia Nora Marlina, Junadhi, Laili Suryati, and Nova Agustina, "Deteksi Dimi Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 11, no. 2, pp. 88–96, May 2022, doi: 10.22146/jnteti.v11i2.3586.
- [12] A. H. Siregar and S. D. Siregar, "Comparison of Logistic Regression and Support Vector Machine Algorithm Performance in Heart Failure Prediction," *Academia Open*, vol. 10, no. 2, Jul. 2025, doi: 10.21070/acopen.10.2025.11682.
- [13] Amrin, Rudianto, and Sismadi, "JITE (Journal of Informatics and Telecommunication Engineering) Data Mining with Logistic Regression and Support Vector Machine for Hepatitis Disease Diagnosis," *JITE*, vol. 8, no. 2, 2025, doi: 10.31289/jite.v8i2.13218.
- [14] 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," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [15] A. F. Al Farizi and Y. Sibaroni, "Implementation of BiLSTM and IndoBERT for Sentiment Analysis of TikTok Reviews," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 1, pp. 96–106, Jan. 2025, doi: 10.29100/jipi.v10i1.5815.
- [16] A. Yoga Pratama, G. Ananda Sanjaya, N. Khairunisa Lubis, and M. Rangga Aditya, "Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial," vol. 9, p. 2025, doi: 10.47002/metik.v9i1.1055.
- [17] N. Hadi and D. Sugiarjo, "Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 1, pp. 37–49, Jan. 2025, doi: 10.30591/jpit.v10i1.7106.
- [18] L. Enjelia, Y. Cahyana, Rahmat, and D. Wahiddin, "Comparison of K-Nearest Neighbors and Naive Bayes Classifier Algorithms in Sentiment Analysis of 2024 Election in Twitter (X)," *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 946–954, Jun. 2025, doi: 10.30871/jaic.v9i3.9593.
- [19] J. Setyanto and T. B. Sasongko, "Sentiment Analysis of Sirekap Application Users Using the Support Vector Machine Algorithm," *Journal of Applied Informatics and Computing*, vol. 8, no. 1, pp. 71–76, Jul. 2024, doi: 10.30871/jaic.v8i1.7772.