

Analisis Sentimen dalam Kasus Korupsi PT. Pertamina menggunakan Metode indoBERT dan RCNN

Sentiment Analysis on the PT Pertamina Corruption Case using IndoBERT and RCNN Methods

¹Wildan Jaya Kusoema*, ²Ichsan Ibrahim

^{1,2}Program Studi Teknik Informatika, STMIK Indonesia Mandiri Bandung

^{1,2}Jl. Belitung No.7, Merdeka, Kec. Sumur Bandung, Kota Bandung, Jawa Barat, Indonesia

*e-mail: wildanj14@gmail.com

(received: 5 June 2025, revised: 17 June 2025, accepted: 18 June 2025)

Abstrak

Penelitian ini bertujuan untuk mengevaluasi performa model hybrid IndoBERT-RCNN dalam klasifikasi sentimen publik terhadap kasus korupsi PT. Pertamina, dengan menekankan pengaruh kombinasi hyperparameter terhadap akurasi model. Dataset yang digunakan terdiri atas 10.078 komentar dari YouTube yang dikumpulkan melalui YouTube Data API, kemudian diproses melalui tahap preprocessing, pelabelan otomatis menggunakan model RoBERTa berbahasa Indonesia, serta penyeimbangan distribusi kelas melalui undersampling dan augmentasi berbasis contextual embedding IndoBERT. Arsitektur model menggabungkan IndoBERT sebagai feature extractor dan RCNN sebagai classifier, serta diuji dengan berbagai kombinasi learning rate dan batch size. Hasil eksperimen menunjukkan bahwa konfigurasi optimal dicapai pada learning rate $2e-5$ dan batch size 16, dengan akurasi 84% dan F1-score 83%. Meskipun model secara umum mampu mengklasifikasikan komentar negatif dengan sangat baik, akurasi pada kelas netral dan positif cenderung lebih rendah karena tumpang tindih fitur semantik dan ambiguitas dalam ekspresi komentar. Penelitian ini memberikan kontribusi spesifik dalam pengembangan analisis sentimen berbahasa Indonesia dengan: (1) mengintegrasikan arsitektur IndoBERT-RCNN untuk isu sosial-politik, (2) mengevaluasi kombinasi hyperparameter secara sistematis pada data opini publik tiga kelas, dan (3) memanfaatkan komentar YouTube sebagai sumber data informal yang relevan. Temuan ini berpotensi diterapkan dalam sistem pemantauan opini publik digital terhadap isu strategis nasional secara real-time.

Kata kunci: Analisis sentimen, IndoBERT, RCNN, hyperparameter, korupsi, pembelajaran mendalam, YouTube, opini publik

Abstract

This study aims to evaluate the performance of a hybrid IndoBERT-RCNN model in classifying public sentiment toward the PT Pertamina corruption case, with a focus on how different hyperparameter combinations affect model accuracy. The dataset consists of 10,078 YouTube comments collected via the YouTube Data API, which were then preprocessed, automatically labeled using an Indonesian-language RoBERTa model, and balanced through class distribution techniques including undersampling and contextual embedding-based augmentation with IndoBERT. The model architecture integrates IndoBERT as a feature extractor and RCNN as the classifier, and was tested using various combinations of learning rates and batch sizes. Experimental results show that the optimal configuration was achieved with a learning rate of $2e-5$ and a batch size of 16, resulting in an accuracy of 84% and an F1-score of 83%. While the model demonstrated strong performance in classifying negative comments, accuracy for neutral and positive classes was relatively lower due to semantic overlap and ambiguity in user expressions. This study contributes to Indonesian-language sentiment analysis by: 1. Integrating the IndoBERT-RCNN architecture for social-political issues, 2. Systematically evaluating hyperparameter combinations for

<http://sistemasi.ftik.unisi.ac.id>

three-class public opinion data, and 3. Utilizing YouTube comments as a relevant source of informal public discourse. The findings have potential applications in real-time digital public opinion monitoring systems for strategic national issues.

Keywords: Sentiment analysis, IndoBERT, RCNN, hyperparameters, corruption, deep learning, YouTube, public opinion

1. Pendahuluan

Kasus korupsi yang melibatkan PT. Pertamina mencuat ke permukaan publik karena nilai kerugian negara yang sangat signifikan, diperkirakan mencapai Rp 193,7 triliun. Kasus ini tidak hanya memengaruhi persepsi masyarakat terhadap perusahaan milik negara, tetapi juga berpotensi mengganggu kepercayaan publik terhadap pemerintahan secara menyeluruh [1] [2]. Dalam era digital, opini publik mengenai isu-isu strategis seperti korupsi semakin banyak diekspresikan melalui media sosial dan platform berbagi video seperti YouTube. Fenomena ini menciptakan ekosistem data yang kaya akan emosi, persepsi, dan kritik dari publik terhadap peristiwa-peristiwa nasional. Salah satu video terkait kasus Pertamina, yang memiliki lebih dari 1,5 juta penayangan dan 10.078 komentar, merepresentasikan ekosistem data publik yang sangat relevan untuk diteliti.

Komentar-komentar yang muncul di platform seperti YouTube mencerminkan spektrum emosi masyarakat terhadap isu yang sedang berkembang. Namun, data semacam ini cenderung tidak terstruktur, mengandung ragam ekspresi informal, serta mengandung konteks ambigu yang menyulitkan proses klasifikasi secara otomatis. Tantangan ini diperparah oleh adanya variasi bahasa yang sangat luas, termasuk singkatan, emotikon, serta struktur kalimat yang tidak mengikuti kaidah formal. Oleh karena itu, pemrosesan komentar publik secara sistematis membutuhkan pendekatan teknologi yang mampu menangkap makna kontekstual secara mendalam. Pendekatan Natural Language Processing (NLP) modern, khususnya model berbasis deep learning seperti transformer, menawarkan solusi yang potensial dalam menjawab tantangan tersebut.

Beberapa dari penelitian sebelumnya menggunakan algoritma machine learning, seperti support vector machine [3], [4], naïve bayes [5], [6], dan juga random forest [5], [7]. Meskipun metode tersebut cukup berhasil untuk dataset yang lebih sederhana, keterbatasan utamanya adalah ketidakmampuannya dalam memahami konteks kalimat secara menyeluruh. Analisis sentimen terhadap komentar sosial yang kompleks membutuhkan model yang tidak hanya memproses kata secara individual, tetapi juga memahami relasi dan makna kontekstual antar kata. Oleh karena itu, diperlukan pendekatan NLP yang lebih canggih, yang tidak hanya bergantung pada frekuensi kata atau fitur statis, melainkan juga memahami makna secara dinamis melalui representasi kontekstual.

Penelitian ini mengusulkan kombinasi model *IndoBERT*—sebuah model berbahasa Indonesia berbasis transformer yang telah ditraining menggunakan lebih dari 4 miliar token bahasa Indonesia—dengan *Recurrent Convolutional Neural Network (RCNN)* sebagai classifier. *IndoBERT* digunakan untuk melakukan ekstraksi fitur dan augmentasi data melalui fine-tuning model “*indobenchmark/indobert-base-p2*”, sedangkan RCNN dimanfaatkan untuk menangkap pola spasial dan sekuensial dalam data teks [8]. Dengan pendekatan ini, diharapkan model dapat lebih efektif dalam menangani data komentar sosial yang tidak terstruktur dan penuh dengan nuansa emosional. Penelitian ini juga secara eksplisit mengeksplorasi pengaruh kombinasi hyperparameter seperti *learning rate* dan *batch size*, untuk mengidentifikasi konfigurasi pelatihan optimal dalam konteks analisis sentimen bahasa Indonesia.

Permasalahan yang dikaji dalam penelitian ini mencakup dua aspek utama, yaitu bagaimana mengklasifikasikan sentimen masyarakat terhadap kasus korupsi Pertamina (positif, negatif, netral), serta bagaimana menentukan kombinasi terbaik dari parameter *learning rate* (5e-5, 3e-5, 2e-5) dan *batch size* (16, 32) dalam proses pelatihan model [9]. Dengan pendekatan ini, diharapkan model dapat lebih efektif dalam menangani data komentar sosial yang tidak terstruktur dan penuh dengan nuansa emosional. Penelitian ini juga secara eksplisit mengeksplorasi pengaruh kombinasi hyperparameter

seperti *learning rate* dan *batch size*, untuk mengidentifikasi konfigurasi pelatihan optimal dalam konteks analisis sentimen bahasa Indonesia.

2. Tinjauan literatur

IndoBERT telah menjadi salah satu model bahasa berbasis transformer yang dominan digunakan dalam penelitian NLP berbahasa Indonesia. Model ini dilatih dengan lebih dari 4 miliar token dan memperlihatkan kinerja yang kompetitif dalam banyak tugas klasifikasi teks. Dalam konteks sentimen, beberapa studi telah mengevaluasi keefektifan model ini, baik secara tunggal maupun dalam arsitektur hybrid.

Penelitian sebelumnya mengombinasikan IndoBERT dengan RCNN untuk klasifikasi sentimen ulasan produk Indonesia dan membuktikan bahwa pendekatan tersebut memberikan F1-score yang lebih tinggi dibandingkan CNN atau LSTM konvensional [8]. Relevansi studi ini terhadap penelitian kami sangat kuat, karena menunjukkan bahwa RCNN dapat mengeksplorasi representasi fitur IndoBERT secara lebih efisien dari IndoBERT. Namun, kelemahan penelitian tersebut adalah tidak adanya analisis terhadap pengaruh hyperparameter seperti *learning rate* dan *batch size*, yang justru menjadi fokus utama dalam penelitian kami.

Penelitian lain meneliti IndoBERT dalam klasifikasi ulasan hotel, dan menemukan bahwa model ini efektif dalam menangkap konteks kalimat yang ambigu [10]. Kelebihan studi ini adalah menunjukkan kemampuan IndoBERT dalam konteks multi-topik dan bahasa informal, mirip dengan komentar di YouTube. Namun, model hanya diuji pada satu konfigurasi tetap tanpa eksplorasi hyperparameter atau kombinasi dengan arsitektur lain, sehingga tidak memberikan wawasan tentang optimalisasi performa model.

Sampai saat ini, kajian terhadap optimasi hyperparameter seperti *learning rate* dan *epoch* memang telah dilakukan dalam studi sebelumnya [9], namun konteks data yang digunakan masih terbatas pada ulasan aplikasi komersial yang bersifat formal (seperti MyTelkomsel). Berbeda dengan itu, penelitian ini mengangkat opini masyarakat yang lebih kompleks dan emosional, yaitu komentar publik YouTube dalam kasus korupsi PT Pertamina. Selain itu, penelitian ini juga mengevaluasi peran RCNN sebagai classifier serta menguji kombinasi *batch size*, aspek yang belum dibahas dalam studi tersebut.

Penelitian lain menggunakan IndoBERT untuk menganalisis komentar “*toxic*” di media sosial dan berhasil menangani data yang tidak terstruktur [11]. Kelebihan studi ini adalah kemampuan generalisasi pada data sosial yang kompleks. Namun, topiknya lebih fokus pada deteksi toksisitas, bukan klasifikasi sentimen dengan nuansa netral-positif-negatif, serta tidak melakukan eksplorasi arsitektur hybrid.

Penelitian lainnya mengembangkan arsitektur hybrid IndoBERT-RCNN-Attention untuk klasifikasi emosi, dan hasilnya menunjukkan peningkatan sensitivitas terhadap makna emosional [12]. Studi ini memberikan justifikasi penggunaan RCNN sebagai classifier karena mampu menangkap dependensi spasial dan temporal dalam fitur dari IndoBERT. Namun, penelitian tersebut menggunakan korpus bahasa Inggris dan domain non-sosial-politik, sehingga aplikasinya pada isu-isu seperti korupsi di Indonesia masih belum teruji.

Berdasarkan ulasan literatur yang telah dikaji, dapat disimpulkan bahwa sebagian besar penelitian terdahulu lebih menekankan efektivitas IndoBERT sebagai feature extractor dalam tugas-tugas klasifikasi teks. Namun, eksplorasi terhadap peran RCNN sebagai classifier yang dikombinasikan dengan IndoBERT masih sangat terbatas. Selain itu, belum ditemukan penelitian yang secara sistematis mengevaluasi kombinasi hyperparameter seperti *learning rate* dan *batch size* dalam konteks klasifikasi sentimen terhadap opini masyarakat pada isu-isu besar berskala nasional, khususnya yang bersumber dari komentar YouTube yang bersifat tidak terstruktur dan informal.

Dengan demikian, penelitian ini mengambil posisi unik dengan mengisi dua celah tersebut secara bersamaan, yaitu melalui evaluasi kombinasi arsitektur IndoBERT-RCNN dan optimasi hyperparameter secara eksperimental. Pendekatan ini tidak hanya memperluas metode teknis yang

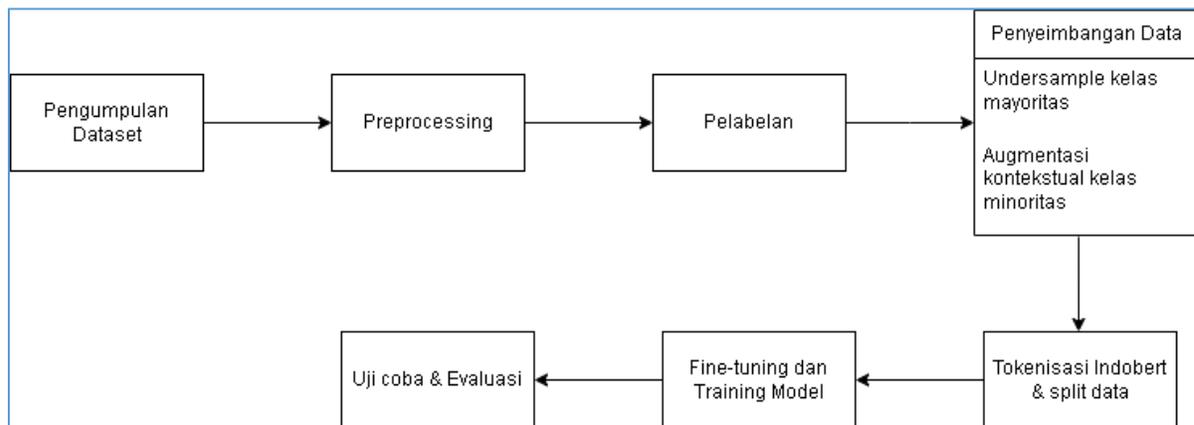
telah ada dalam pemrosesan bahasa alami berbahasa Indonesia, tetapi juga memperkuat dimensi sosial-politik dalam penerapannya, khususnya dalam memahami dinamika opini publik terhadap isu strategis nasional.

3. Metode Penelitian

Penelitian ini menerapkan pendekatan eksperimen kuantitatif untuk menganalisis sentimen masyarakat terhadap kasus korupsi PT. Pertamina berdasarkan komentar YouTube. Model yang dibangun menggunakan pendekatan *hybrid deep learning* dengan kombinasi IndoBERT sebagai *feature extractor* dan RCNN sebagai *classifier*. Penelitian ini mengkaji pengaruh kombinasi hyperparameter terhadap performa klasifikasi sentimen tiga kelas.

Gambar 1 menggambarkan alur metodologi penelitian yang terdiri dari tahapan-tahapan utama yang dilakukan secara berurutan. Dimulai dari tahap *pengumpulan data*, di mana komentar-komentar publik dari platform YouTube dikumpulkan menggunakan API resmi. Selanjutnya dilakukan proses *preprocessing*, yaitu pembersihan dan normalisasi data untuk mengatasi noise linguistik. Setelah itu, data diberi label sentimen secara otomatis menggunakan model RoBERTa bahasa Indonesia, kemudian dilakukan penyeimbangan dataset melalui strategi *undersampling* dan *contextual word embedding augmentation*.

Setelah data siap, model hybrid IndoBERT-RCNN dilatih dengan berbagai konfigurasi hyperparameter untuk mengevaluasi performa klasifikasi sentimen terhadap tiga kelas (positif, netral, negatif). Evaluasi model dilakukan berdasarkan metrik akurasi, presisi, recall, dan F1-score. Gambar ini memberikan visualisasi menyeluruh atas pendekatan sistematis penelitian dan menunjukkan keterkaitan antar tahap untuk mendukung replikasi atau pengembangan lanjutan.



Gambar 1. Alur penelitian

3.1. Pengumpulan Data

Data komentar diperoleh melalui YouTube Data API dari sebuah video yang membahas kasus korupsi PT. Pertamina. Komentar dikumpulkan dari tanggal awal publikasi video hingga 25 Mei 2024. Total komentar yang diperoleh adalah 10.078, yang kemudian disaring dan diproses lebih lanjut untuk mendapatkan data bersih yang layak untuk pelab dan pelatihan model.

Tabel 1 menyajikan contoh representatif dari data komentar yang dikumpulkan melalui YouTube Data API. Setiap entri terdiri atas kolom komentar, nama pengguna, jumlah suka (like), dan waktu publikasi. Komentar-komentar ini mencerminkan reaksi publik terhadap isu korupsi PT. Pertamina dan mencakup berbagai ekspresi emosional serta opini spontan yang muncul di ruang digital. Data dalam tabel menunjukkan kecenderungan penggunaan bahasa informal, ekspresi langsung, dan nada kritik yang menjadi karakteristik khas komentar netizen dalam kasus ini. Contoh seperti "SAHKAN UU Perampasan aset DAN HUKUM MATI KORUPTOR"

menggambarkan nada negatif dan marah, sementara komentar lain lebih bersifat sarkastik atau sinis. Informasi ini membantu menggambarkan tantangan awal yang dihadapi pada tahap preprocessing dan pelabelan, terutama dalam hal noise linguistik, ketidakterstruktur data, dan keragaman semantik.

Tabel 1. Sample dataset

| Comment | Author | Like Count | Published At |
|--|--------------------|------------|----------------------|
| Pelaku nya aja dikawal kayak presiden dikasih naik mobil baik baik tapi mngkn kalau maling ayam atau hp sdh dihajar habis-habisan sama petugas kan aneh. | @br4dlyj648 | 0 | 2025-04-10T12:56:26Z |
| Dpr bisa apa bisanya duduk duduk make jas. Dan nutup nutupi koruptor hadeh payah | @henisetiawati9895 | 0 | 2025-03-08T08:58:45Z |
| SAHKAN UU Perampasan asett DAN HUKUM MATI KORUPTOR | @boy_azi | 0 | 2025-03-10T10:41:26Z |

3.2. Preprocessing

Preprocessing bertujuan untuk menyiapkan data yang belum terstruktur dan memiliki noise menjadi terstruktur dan bersih agar dapat dianalisa lebih lanjut[13]. Preprocessing sendiri memiliki beberapa tahapan, antara lain:

1. *Case folding*, bertujuan untuk mengubah seluruh teks menjadi huruf kecil.
2. *Cleaning data*, bertujuan untuk membersihkan data dari konteks yang tidak relevan, termasuk url, karakter non-alfanumerik, hastag, emot, dan spasi berlebih.
3. *Normalization*, bertujuan untuk mengganti bahasa yang tidak baku menjadi bahasa baku.
4. *Stopwords removal*, bertujuan untuk menghapus kata-kata yang umum dan tidak bermakna, seperti “dan”, “di”, “yang”.

3.3. Pelabelan Sentimen

Setelah melalui tahapan pembersihan, data kemudian diberi label sentimen menggunakan model pre-trained "w11wo/indonesian-roberta-base-sentiment-classifier", yang dikembangkan untuk klasifikasi sentimen teks berbahasa Indonesia [14]. Model ini secara otomatis mengkategorikan komentar ke dalam tiga kelas utama, yaitu positif, netral, dan negatif.

Untuk mengatasi permasalahan ketidakseimbangan distribusi kelas yang umum terjadi dalam data opini publik, dilakukan dua strategi penyesuaian data. Pertama, dilakukan undersampling pada kelas mayoritas (negatif), sehingga jumlahnya diseimbangkan dengan kelas lain. Kedua, untuk kelas minoritas seperti netral dan positif, diterapkan augmentasi data menggunakan teknik *Contextual Word Embedding Augmentation* berbasis IndoBERT. Teknik ini menghasilkan variasi kalimat baru dengan mengganti kata-kata dalam komentar menggunakan embedding kontekstual, tanpa mengubah makna inti, sehingga meningkatkan keragaman dan jumlah sampel untuk kelas yang kurang terwakili[15].

Setelah proses augmentasi dan penyeimbangan selesai, seluruh dataset diacak ulang untuk memastikan distribusi label yang proporsional dan representatif dalam proses pelatihan model klasifikasi sentimen.

3.4. Arsitektur model

Model yang digunakan merupakan implementasi dari arsitektur IndoBERT-RCNN, yaitu integrasi antara representasi kontekstual dari model IndoBERT dan mekanisme feature extraction serta sequence modeling dari RCNN. Pendekatan ini menghindari ketergantungan pada token [CLS] saja dan justru memanfaatkan seluruh output token BERT, yang terbukti meningkatkan kinerja klasifikasi secara signifikan dalam tugas analisis sentimen bahasa Indonesia [8]. Arsitektur model terdiri atas:

1. IndoBERT (indobenchmark/indobert-base-p2) sebagai *feature extractor*, menghasilkan representasi kontekstual dari teks.
2. Conv1D Layer: Layer konvolusional satu dimensi dengan 128 filter untuk mengekstraksi fitur spasial dari representasi token.
3. BiLSTM Layer: LSTM dua arah dengan 64 unit untuk menangkap pola sekuensial dalam teks.
4. Fully Connected Layer: Layer terakhir dilengkapi dengan fungsi aktivasi **softmax** yang mengubah skor logit menjadi probabilitas untuk masing-masing dari tiga kelas sentimen: positif, netral, dan negatif.

Data tokenisasi dilakukan menggunakan tokenizer IndoBERT dengan panjang maksimum 128 token. Dataset dibagi menjadi tiga bagian: 80% untuk pelatihan, 10% untuk validasi, dan 10% untuk pengujian, menggunakan stratifikasi label untuk menjaga distribusi kelas.

3.5. Evaluasi Model

Model mengevaluasi performa model terhadap enam kombinasi hyperparameter, yaitu tiga nilai *learning rate* ($5e-5$, $3e-5$, $2e-5$) dan dua ukuran *batch size* (16 dan 32). Evaluasi dilakukan dengan metrik accuracy, precision, recall, dan f1-score. Visualisasi hasil dilakukan melalui grafik *learning curve*, *accuracy curve*, dan *confusion matrix*. Hasil dari seluruh konfigurasi dibandingkan untuk menentukan kombinasi parameter terbaik [14].

4. Hasil dan Pembahasan

Penelitian ini menyajikan hasil evaluasi dari model IndoBERT-RCNN dalam mengklasifikasikan sentimen publik terhadap kasus korupsi PT. Pertamina berdasarkan komentar YouTube. Pembahasan dimulai dari deskripsi dataset yang digunakan, proses pelabelan dan penyeimbangan data, hingga tahapan tokenisasi dan pelatihan model dengan berbagai konfigurasi hyperparameter. Selanjutnya, hasil evaluasi model ditampilkan secara kuantitatif melalui metrik performa seperti akurasi, presisi, recall, dan F1-score, disertai visualisasi grafik dan confusion matrix.

4.1. Dataset

Data yang digunakan dalam penelitian ini diperoleh melalui *YouTube Data API* dari sebuah video berjudul “*Tanggapi Korupsi di Pertamina, Ini Kata Prabowo*” yang secara langsung membahas isu korupsi pada PT. Pertamina. Video tersebut memiliki tingkat keterlibatan tinggi, yaitu lebih dari 1,5 juta penayangan dan mengambil setidaknya 10.078 komentar untuk pengujian data.

Sebelum memasuki pelatihan model, *preprocessing* diperlukan untuk membersihkan data menyaring komentar-komentar yang tidak relevan. Data seperti promosi, spam, dan komentar diluar konteks akan dihapus dari dataset. Kemudian data dibersihkan dari noise yang akan mengganggu proses pelatihan.

Tabel 2 merupakan perbandingan dataset sebelum dan sesudah preprocessing. Data tidak hanya dibersihkan pada proses ini, namun juga teks yang menggunakan singkatan atau kata tidak baku kemudian dinormalisasikan. Lalu kata sambung seperti “dan”, “di”, “yang”, dihapus dari data.

Tabel 2. Perbandingan sebelum dan setelah *preprocessing*

| Sebelum <i>Preprocessing</i> | Setelah <i>Preprocessing</i> |
|--|--|
| Buktikan pak jangan cm omon2 saja. Buat rakyat percaya | buktikan pak jangan omong kosong buat rakyat percaya |
| ambil semua kekayaan,, miskin kan para korupsi | ambil semua kekayaan miskin kan korupsi |

Setelah tahap pembersihan, jumlah komentar yang layak dianalisis tersisa 9.482 komentar. Data kemudian dilabeli secara otomatis menggunakan model pra-latih “*w11wo/indonesian-roberta-base-sentiment-classifier*” berbasis RoBERTa untuk teks berbahasa Indonesia. Pelabelan dilakukan terhadap tiga kelas sentimen: *positif*, *netral*, dan *negatif*.

Model pelabel otomatis yang digunakan memiliki performa validasi yang tinggi, sebagaimana dilaporkan oleh pengembang resminya, yaitu mencapai akurasi sebesar 94,36%, F1-score 92,42%, precision 92,64%, dan recall 92,20% [16]. Meskipun angka ini menunjukkan keandalan model dalam klasifikasi sentimen, tetap terdapat kemungkinan terjadinya bias atau kesalahan pelabelan, terutama terhadap komentar dengan konteks ambigu atau ekspresi emosional yang kompleks. Karena seluruh proses pelabelan dilakukan secara otomatis dan tidak melalui verifikasi manual, potensi adanya label yang tidak sesuai konteks tetap menjadi keterbatasan yang perlu diperhitungkan dalam mengevaluasi validitas hasil pelatihan model IndoBERT-RCNN.

Data yang sudah diberikan label kemudian diseimbangkan dengan strategi undersampling kelas mayoritas menjadi 3000, sedangkan untuk kelas minoritas seperti netral dan positif dilakukan augmentasi data. Teknik augmentasi yang digunakan adalah “*Contextual Word Embedding Augmentation*” berbasis model IndoBERT, yaitu proses menghasilkan variasi baru dari kalimat asli dengan mengganti kata-kata tertentu menggunakan konteks vektor embedding yang diperoleh dari IndoBERT.



Gambar 2. Distribusi sentimen sebelum dan sesudah penyeimbangan data

Pada gambar 2 menunjukkan hasil perbandingan distribusi data sebelum dan sesudah penyeimbangan, yang semula menunjukkan dominasi besar pada kelas negatif dan ketimpangan signifikan dengan kelas lain. Setelah penyesuaian, distribusi menjadi lebih proporsional di ketiga kelas, yang bertujuan untuk mengurangi potensi bias model dalam proses pelatihan serta meningkatkan akurasi prediksi pada kelas netral dan positif yang sebelumnya kurang terwakili.

Sementara itu, tabel 3 menunjukkan jumlah aktual sebelum maupun sesudah penyeimbangan data, yang mana kelas *negative* memiliki 3000 data, kelas *neutral* memiliki 3456 data, dan kelas *positive* memiliki 3528 data.

Tabel 3. Perbandingan sebelum dan setelah penyeimbangan data

| Sentimen | Sebelum Penyeimbangan Data | Setelah Penyeimbangan Data |
|----------|----------------------------|----------------------------|
| Negative | 6578 | 3000 |
| Neutral | 1728 | 3456 |
| Positive | 1176 | 3528 |

4.2. Tokenisasi IndoBERT

Tokenisasi dilakukan menggunakan model (indobenchmark/indobert-base-p2), yaitu varian BERT yang sudah di-pretrain dengan korpus Bahasa Indonesia. Tokenizer ini bertugas memecah teks menjadi unit terkecil (token) dan mengubahnya ke dalam ID vektor numerik yang dapat diproses oleh model BERT [17]. Distribusi panjang token dalam data training menunjukkan sebagian besar teks memiliki panjang di bawah 50 token, yang menunjukkan efisiensi model terhadap teks pendek menengah.

Tabel 4. Hasil tokenisasi

| Teks Asli | Tokens | Encode |
|---|---|--|
| allah sehatkan panjangkan umurnya bapak prabowo allah | ['allah', 'sehat', '##kan', 'panjang', '##kan', 'umurnya', 'bapak', 'prabowo', 'allah'] | [763, 1705, 32, 1422, 32, 14885, 2811, 11283, 763] |

Tabel 4 menjelaskan bahwa token seperti asetnya dipecah menjadi dua bagian, yaitu “asset” dan “##nya”. Ini mencerminkan kemampuan morfologis IndoBERT untuk memahami bentuk turunan dari kata dasar. Untuk sub-token “##nya” memiliki ID sendiri, terpisah dari kata sebelumnya, menandakan bahwa IndoBERT memperlakukan kata dasar dan afiks sebagai unit informasi terpisah. Tokenisasi ini membuat model tetap mampu memahami arti dari *kata turunan atau frasa idiomatik*, meskipun tidak semua bentuk kata muncul dalam pretraining.

4.3. Evaluasi model

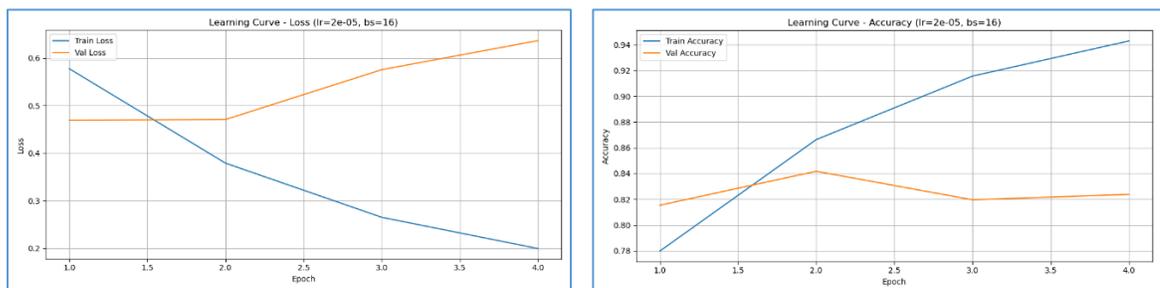
Pada pelatihan model IndoBERT dan RCNN, dilakukan pengujian dengan konfigurasi hyperparameter dengan learning rate 5e-5, 3e-5, dan 2e-5; dengan batch size 16, dan 32. Tabel 5 menunjukkan hasil evaluasi berdasarkan *accuracy*, *precision*, *recall*, dan *f1-score*.

Tabel 5 Evaluasi Hyperparameter

| LRate | Batch | Epoch | Acc | Prec | Recall | F1 |
|-------|-------|-------|--------|--------|--------|--------|
| 5e-05 | 16 | 1 | 0.7869 | 0.7900 | 0.7869 | 0.7761 |
| | | 2 | 0.8059 | 0.8087 | 0.8059 | 0.8037 |
| | | 3 | 0.8038 | 0.8190 | 0.8038 | 0.7878 |

| | | | | | | |
|-------|----|---|---------------|---------------|---------------|---------------|
| | | 4 | 0.7700 | 0.7988 | 0.7700 | 0.7734 |
| | | 5 | 0.7880 | 0.7793 | 0.7880 | 0.7771 |
| | 32 | 1 | 0.7911 | 0.8183 | 0.7911 | 0.7990 |
| | | 2 | 0.8091 | 0.8067 | 0.8091 | 0.8075 |
| | | 3 | 0.8154 | 0.8089 | 0.8154 | 0.8045 |
| | | 4 | 0.8154 | 0.8068 | 0.8154 | 0.8081 |
| | | 5 | 0.7964 | 0.8022 | 0.7964 | 0.7942 |
| 3e-05 | 16 | 1 | 0.8302 | 0.8250 | 0.8302 | 0.8218 |
| | | 2 | 0.8259 | 0.8191 | 0.8259 | 0.8210 |
| | | 3 | 0.8249 | 0.8170 | 0.8249 | 0.8155 |
| | | 4 | 0.8101 | 0.8016 | 0.8101 | 0.8038 |
| | | 5 | 0.8133 | 0.8037 | 0.8133 | 0.7997 |
| | 32 | 1 | 0.8196 | 0.8125 | 0.8196 | 0.8117 |
| | | 2 | 0.8091 | 0.8102 | 0.8091 | 0.7910 |
| | | 3 | 0.8091 | 0.8102 | 0.8091 | 0.7910 |
| | | 4 | 0.8186 | 0.8109 | 0.8186 | 0.8112 |
| | | 5 | 0.7922 | 0.7974 | 0.7922 | 0.7941 |
| 2e-05 | 16 | 1 | 0.8154 | 0.8088 | 0.8154 | 0.8099 |
| | | 2 | 0.8418 | 0.8372 | 0.8418 | 0.8320 |
| | | 3 | 0.8196 | 0.8174 | 0.8196 | 0.8075 |
| | | 4 | 0.8238 | 0.8250 | 0.8238 | 0.8244 |
| | | 5 | 0.8154 | 0.8129 | 0.8154 | 0.7959 |
| | 32 | 1 | 0.8165 | 0.8213 | 0.8165 | 0.8034 |
| | | 2 | 0.8259 | 0.8200 | 0.8259 | 0.8216 |
| | | 3 | 0.8091 | 0.8112 | 0.8091 | 0.8053 |
| | | 4 | 0.8091 | 0.8112 | 0.8091 | 0.8053 |
| | | 5 | 0.8291 | 0.8220 | 0.8291 | 0.8220 |
| | | 6 | 0.8196 | 0.8163 | 0.8196 | 0.8159 |

Berdasarkan hasil pengujian hyperparameter yang ada pada Tabel 5, model terbaik yang diperoleh terdapat pada konfigurasi learning rate 2e-5 dan batch size 16 dengan accuracy 0.84 dan f1-score sebanyak 0.83. Sedangkan untuk model terendah terdapat pada learning rate 5e-5, dan batch size 16 dengan accuracy 0.77 dan f1-score sebanyak 0.77.

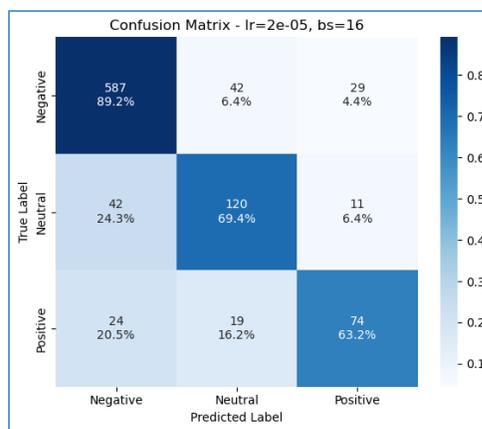


Gambar 3. Learning curve model terbaik

Berdasarkan grafik *learning curve* yang dihasilkan selama proses pelatihan model IndoBERT-RCNN di Gambar 3, terlihat adanya indikasi terjadinya overfitting yang muncul sejak *epoch* ketiga. Grafik pertama menunjukkan tren *loss* pada data pelatihan dan validasi. Nilai *train loss* mengalami penurunan konsisten dari 0.577 pada *epoch* pertama menjadi 0.199 pada *epoch* keempat. Hal ini mengindikasikan bahwa model berhasil mempelajari pola dari data pelatihan dengan baik.

Namun, *val loss* justru menunjukkan pola yang kontras. Setelah stabil pada *epoch* pertama dan kedua di kisaran 0.469, nilai *val loss* meningkat pada *epoch* ketiga dan keempat, masing-masing mencapai 0.575 dan 0.636. Peningkatan ini menunjukkan bahwa performa model terhadap data validasi memburuk, yang merupakan indikasi awal dari overfitting, di mana model terlalu menyesuaikan diri terhadap data pelatihan dan kehilangan kemampuan generalisasi terhadap data yang tidak terlihat sebelumnya.

Grafik accuracy menunjukkan perkembangan akurasi pada data pelatihan dan validasi. *Train accuracy* meningkat secara konsisten dari 0.78 pada *epoch* pertama menjadi 0.943 pada *epoch* keempat. Sebaliknya, *val accuracy* hanya meningkat hingga *epoch* kedua (0.842), kemudian mengalami penurunan pada *epoch* ketiga (0.820) dan sedikit meningkat pada *epoch* keempat (0.824), namun tidak signifikan jika dibandingkan dengan peningkatan pada data pelatihan. Untuk memitigasi overfitting yang diduga disebabkan oleh penggunaan pelabelan otomatis secara menyeluruh, disarankan penerapan strategi semi-supervised learning atau validasi manual parsial guna meningkatkan keandalan label sebagai dasar pembelajaran model.



Gambar 4 Confusion matrix model

Confusion matrix menunjukkan kinerja yang sangat baik pada kelas *Negatif*, dengan 587 dari 658 sampel (sekitar 89,2%) berhasil diklasifikasikan dengan benar. Kesalahan klasifikasi terhadap kelas *Netral* dan *Positif* relatif kecil, masing-masing sebesar 6,4% dan 4,4%.

Sementara performa model pada kelas *Netral* lebih rendah dibandingkan kelas *Negatif*, dengan hanya 69,4% (120 dari 173) sampel yang berhasil diklasifikasikan dengan benar, dan sisanya salah diklasifikasikan sebagai *Negatif* (24,3%) dan *Positif* (6,4%). Meskipun jumlah sampel kelas *Netral* pasca-augmentasi bahkan sedikit lebih besar dari kelas *Negatif*, hasil ini mengindikasikan adanya kesulitan model dalam membedakan ekspresi netral dari negatif, yang kemungkinan besar disebabkan oleh ambiguitas semantik komentar, overlap fitur vektorial dalam embedding IndoBERT, serta potensi kesalahan pelabelan otomatis pada komentar bermakna ganda. Komentar netral seperti “mana buktinya 4 bulan krugat kurget berita tindak lanjutnya” atau “oke gas oke gas ndasmu” bersifat tidak eksplisit secara emosional, sehingga lebih rentan diklasifikasikan ke kelas lain oleh model. Oleh karena itu, tantangan pada kelas *Netral* ini lebih disebabkan oleh kompleksitas semantik dan kualitas label, bukan sekadar kuantitas data.

Model mengalami kesulitan yang lebih besar dalam mengenali kelas *Positif*. Hanya 63,2% (74 dari 117) sampel yang diklasifikasikan dengan benar. Sebanyak 20,5% justru diklasifikasikan sebagai *Negatif*, dan 16,2% sebagai *Netral*. Hal ini menandakan adanya tumpang tindih fitur antara kelas *Positif* dan kelas lainnya, atau representasi kelas *Positif* yang kurang kuat pada proses pelatihan.

5. Kesimpulan

Penelitian ini menunjukkan bahwa kombinasi hyperparameter memiliki peran signifikan terhadap performa model IndoBERT-RCNN dalam klasifikasi sentimen publik terkait kasus korupsi PT. Pertamina. Dengan membandingkan berbagai nilai learning rate dan batch size, konfigurasi optimal ditemukan pada learning rate $2e-5$ dan batch size 16, yang memberikan akurasi tertinggi sebesar 84% dan F1-score mencapai 83%. Temuan ini mengindikasikan bahwa penyesuaian hyperparameter secara tepat dapat meningkatkan efektivitas model deep learning dalam menangkap nuansa emosional dalam teks berbahasa Indonesia.

Model menunjukkan performa unggul dalam mendeteksi sentimen negatif. Namun, klasifikasi terhadap kelas netral dan positif masih menghadapi tantangan akibat tumpang tindih fitur, ambiguitas semantik, serta potensi kesalahan pada proses pelabelan otomatis. Selain itu, analisis learning curve menunjukkan indikasi overfitting sejak epoch ketiga, yang memperkuat pentingnya evaluasi kualitas label serta kebutuhan akan strategi pelabelan yang lebih cermat di masa depan.

Penelitian ini memberikan kontribusi konkret terhadap pengembangan sistem analisis sentimen berbasis bahasa Indonesia melalui:

1. Integrasi pertama arsitektur IndoBERT-RCNN dalam konteks opini publik terhadap isu sosial-politik nasional, yang belum banyak dieksplorasi secara sistematis sebelumnya.
2. Eksperimen komprehensif terhadap kombinasi hyperparameter dalam skenario tiga kelas sentimen (positif, netral, negatif), menghasilkan panduan konfigurasi yang dapat direplikasi.
3. Penerapan augmentasi berbasis contextual embedding IndoBERT, yang membantu mengatasi ketidakseimbangan distribusi kelas secara efektif.
4. Penggunaan data komentar YouTube real-world yang informal dan tidak terstruktur, yang merepresentasikan tantangan analisis opini publik dalam lingkungan sosial digital Indonesia.

Sebagai implikasi praktis, pendekatan ini dapat dimanfaatkan dalam pengembangan sistem pemantauan opini publik real-time oleh pemerintah, media, atau lembaga swadaya masyarakat sebagai alat bantu pengambilan keputusan terhadap isu-isu strategis nasional.

Untuk arah penelitian selanjutnya, beberapa langkah yang disarankan meliputi:

1. Implementasi pelabelan semi-supervised atau validasi manual sebagian untuk meningkatkan kualitas dataset, terutama pada komentar berlabel netral.
2. Eksplorasi arsitektur alternatif berbasis transformer murni, seperti IndoBERT dengan attention layer atau adapter layer.
3. Uji coba model pada platform sosial media lain, seperti TikTok atau X (Twitter), guna menilai generalisasi model terhadap domain berbeda.
4. Integrasi analisis temporal untuk melacak dinamika opini publik dari waktu ke waktu.
5. Pengembangan sistem prototipe real-time sentiment monitor sebagai alat bantu pengawasan persepsi masyarakat terhadap isu nasional penting.

Dengan demikian, penelitian ini tidak hanya memperluas pendekatan teknis yang telah ada, tetapi juga memperkuat relevansi sosial dan praktis dari teknologi pemrosesan bahasa alami dalam konteks Indonesia kontemporer.

Referensi

- [1] A. Mulyadi, "Ada Apa dengan Pertamina? Analisis Hukum terhadap Kasus Korupsi PT. Pertamina Parta Niaga," *Jurnal Anti Korupsi*, Vol. 01, No. 1, pp. 37–48, Jun. 2025, doi: 10.19184/jak.v13i1.40004.
- [2] D. V. Hartanto, F. Navrizal, F. Surya, H. Berlanty, and M. Naufal, "Menegakkan Keadilan dalam Pengelolaan Sumber Daya Negara (Studi Kasus: Korupsi Pengoplosan di Pertamina)," *Inspirasi & Strategi (INSPIRAT): Jurnal Kebijakan Publik & Bisnis*, Vol. 16, No. 01, pp. 1–4, 2025.

- [3] R. D. Pebrianti, “Analisis Sentimen Masyarakat Platform X terhadap Korupsi PT. Pertamina (Persero) menggunakan Metode SVM,” *Jurnal Informatika dan Teknik Elektro Terapan*, Vol. 13, No. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6399.
- [4] M. L. Hermanto, F. Fathoni, O. Ardhillah, and A. Ibrahim, “Analisis Sentimen Masyarakat terhadap DANANTARA di Platform X dengan Metode SVM,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 9, No. 4, pp. 6779–6785, May 2025, doi: 10.36040/jati.v9i4.14189.
- [5] R. I. Agustin, “Komparasi Algoritma Naive Bayes dan SVM untuk Analisis Sentimen Twitter Korupsi Bansos Beras Masa Pandemi,” *Jurnal Informatika dan Teknik Elektro Terapan*, Vol. 12, No. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4020.
- [6] Y. Zaen Vebrian and Kustiyono, “A Sentiment Analysis of Free Meal Plans on Social Media using Naive Bayes Algorithm,” *Jurnal Inovetek Polbeng - Seri Informatika*, Vol. 10, No. 1, pp. 355–366, 2025.
- [7] I. Muhandhis, A. S. Ritonga, W. Putra, and R. Benowo, “Public Sentiment Analysis on TikTok about Tapera Policy using Random Forest Classifier,” *Sistemasi: Jurnal Sistem Informasi*, Vol. 14, No. 1, pp. 354–365, 2025, doi: 10.32520/stmsi.v14i1.4878.
- [8] 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.
- [9] M. N. Zaidan, Y. Sibaroni, and S. S. Prasetyowati, “Learning Rate and Epoch Optimization in the Fine-Tuning Process for IndoBERT’S Performance on Sentiment Analysis of MyTelkomsel App Reviews,” *Jurnal Teknik Informatika (Jutif)*, Vol. 5, No. 5, pp. 1443–1450, Oct. 2024, doi: 10.52436/1.jutif.2024.5.5.2396.
- [10] Y. A. Singgalen, “Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data,” *Article in Journal of Information System Research*, Vol. 6, No. 2, pp. 978–988, 2025, doi: 10.47065/josh.v6i2.6505.
- [11] G. Z. Nabiihah, S. Y. Prasetyo, Z. N. Izdihar, and A. S. Girsang, “BERT Base Model for Toxic Comment Analysis on Indonesian Social Media,” in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 714–721. doi: 10.1016/j.procs.2022.12.188.
- [12] H. Ahmadian, T. F. Abidin, H. Riza, and K. Muchtar, “Hybrid Models for Emotion Classification and Sentiment Analysis in Indonesian Language,” *Applied Computational Intelligence and Soft Computing*, Vol. 2024, No. 1, pp. 1–17, 2024, doi: 10.1155/2024/2826773.
- [13] N. Sholihah, F. F. Abdulloh, and M. Rahardi, “Sentiment Analysis on KPU Performance Post-2024 Election via YouTube Comments using BERT,” *sinkron*, Vol. 8, No. 4, pp. 2222–2232, Oct. 2024, doi: 10.33395/sinkron.v8i4.14040.
- [14] R. Merdiansah, S. Siska, and A. Ali Ridha, “Analisis Sentimen Pengguna X Indonesia terkait Kendaraan Listrik menggunakan IndoBERT,” *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, Vol. 7, No. 1, pp. 221–228, Mar. 2024, doi: 10.55338/jikomsi.v7i1.2895.
- [15] A. Miaschi and F. Dell’orletta, “Contextual and Non-Contextual Word Embeddings: An in-Depth Linguistic Investigation,” in *Proceedings of the 5th Workshop on Representation Learning for NLP*, Spandana Gella, Johannes Welbl, Marek Rei, Fabio Petroni, Patrick Lewis, Emma Strubell, Minjoon Seo, and Minjoon Seo, Eds., Association for Computational Linguistics, Jul. 2020, pp. 110–119. doi: 10.18653/v1/2020.repl4nlp-1.15.
- [16] W. Wongso, “Indonesian-Roberta-Base-Sentiment-Classifer (Revision e402e46),” 2023, *Hugging Face*: e402e46. Accessed: Jun. 04, 2025. [Online]. Available: <https://huggingface.co/w11wo/indonesian-roberta-base-sentiment-classifier>
- [17] W. A. Hidayat and V. R. S. Nastiti, “Perbandingan Kinerja Pre-Trained IndoBERT-Base dan IndoBERT-Lite pada Klasifikasi Sentimen Ulasan Tiktok Tokopedia Seller Center dengan model IndoBERT,” *JSiI (Jurnal Sistem Informasi)*, Vol. 11, No. 2, pp. 13–20, Sep. 2024, doi: 10.30656/jsii.v11i2.9168.