

Fine-Tuning IndoBERT Untuk Analisis Sentimen Berita Saham Berbahasa Indonesia Dengan *Hyperparameter Optimization*

Samsul Alam¹, Zeldy Suryady², Adhi Kusnadi³

^{1,2,3} Magister Informatika, Universitas Nusa Putra

Jl. Raya Cibolang No.21 Cisaat, Sukabumi, Jawa Barat, Indonesia

samsul.alam@nusaputra.ac.id

Abstrak

Penelitian ini bertujuan untuk meningkatkan kinerja analisis sentimen pada berita saham berbahasa Indonesia menggunakan model berbasis *transformer*. Pesatnya pertumbuhan pasar modal Indonesia menyebabkan meningkatnya *volume* informasi dalam bentuk berita, sehingga diperlukan sistem analisis sentimen otomatis yang akurat untuk mendukung pengambilan keputusan investasi berbasis data. Metode yang digunakan adalah *fine-tuning* model IndoBERT dengan pendekatan *hyperparameter optimization* menggunakan *framework* Optuna berbasis *Bayesian Optimization*. Dataset yang digunakan terdiri dari 23.108 artikel berita saham yang diklasifikasikan ke dalam tiga kelas sentimen, yaitu positif, netral, dan negatif. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, *Macro-F1*, *confusion matrix*, serta ROC-AUC dengan pendekatan *one-vs-rest* untuk klasifikasi multi-kelas. Hasil penelitian menunjukkan bahwa model IndoBERT Base Uncased dengan konfigurasi *hyperparameter* optimal memberikan performa terbaik dengan *accuracy* sebesar 0,8269 dan *Macro-F1* sebesar 0,7816. Penerapan *hyperparameter optimization* terbukti mampu meningkatkan performa model secara signifikan dibandingkan *baseline* tanpa optimasi. Penelitian ini berkontribusi dalam pengembangan analisis sentimen berbasis bahasa Indonesia pada domain finansial serta menyediakan pendekatan yang efektif untuk meningkatkan performa model melalui optimasi *hyperparameter*.

Kata kunci: Analisis Sentimen, IndoBERT, *Fine-Tuning*, *Hyperparameter Optimization*, Berita Saham.

Abstract

This study aims to improve sentiment analysis performance on Indonesian stock news using transformer-based models. The rapid growth of the Indonesian capital market has led to an increase in financial news information, creating the need for accurate automated sentiment analysis systems to support data-driven investment decisions. The proposed method applies fine-tuning on the IndoBERT model with hyperparameter optimization using Bayesian Optimization through the Optuna framework. The dataset consists of 23,108 Indonesian stock news articles classified into three sentiment classes: positive, neutral, and negative. Model evaluation is conducted using accuracy, precision, recall, F1-score, Macro-F1, confusion matrix, and ROC-AUC with a one-vs-rest approach for multi-class classification. The results indicate that IndoBERT-Base-Uncased with optimal hyperparameter configuration achieves the best performance, with an accuracy of 0.8269 and an F1-score of 0.7816. The application of hyperparameter optimization significantly improves model performance compared to the baseline. This study contributes to the advancement of Indonesian-language sentiment analysis in the financial domain and provides an effective approach to improving model performance through hyperparameter optimization.

Keywords: Sentiment Analysis, IndoBERT, *Fine-Tuning*, *Hyperparameter Optimization*, Stock News

I. PENDAHULUAN

Pasar modal Indonesia mengalami pertumbuhan yang signifikan dalam beberapa tahun terakhir,

ditandai dengan meningkatnya jumlah investor dan *volume* transaksi. Pertumbuhan ini diikuti dengan meningkatnya produksi informasi finansial, khususnya dalam bentuk berita saham yang tersebar

melalui berbagai *platform* media daring. Informasi tersebut memiliki pengaruh terhadap persepsi investor dan dapat memengaruhi pergerakan harga saham, sehingga analisis sentimen terhadap berita menjadi aspek penting dalam mendukung pengambilan keputusan investasi berbasis data [1], [2].

Seiring dengan perkembangan teknologi, metode analisis sentimen berbasis *Natural Language Processing* (NLP) telah banyak digunakan untuk mengekstraksi opini dari teks [3], [4]. Model berbasis *transformer*, seperti BERT dan turunannya, menunjukkan performa yang unggul dalam berbagai tugas NLP, termasuk klasifikasi teks [5], [6]. Dalam konteks bahasa Indonesia, IndoBERT sebagai model *pre-trained* telah menunjukkan kinerja yang baik pada berbagai tugas pemrosesan bahasa alami [7]. Hal ini menunjukkan bahwa pendekatan berbasis *transformer* memiliki potensi besar untuk diterapkan pada analisis sentimen teks berbahasa Indonesia, termasuk pada domain finansial.

Beberapa penelitian sebelumnya telah mengkaji analisis sentimen menggunakan model berbasis *transformer*. Studi oleh Liu et al [4] menunjukkan bahwa model FinBERT mampu meningkatkan performa analisis teks finansial melalui pendekatan *pre-trained language model*. Sementara itu, penelitian oleh Simanjuntak et al [8] menunjukkan bahwa optimasi *hyperparameter* pada model IndoBERT dapat meningkatkan performa secara signifikan pada tugas deteksi berita palsu. Selain itu, Maharani et al [9] mengembangkan model bahasa khusus domain finansial untuk meningkatkan pemahaman konteks pada teks keuangan berbahasa Indonesia. Penelitian lain oleh Anderies et al [10] juga menunjukkan bahwa pendekatan berbasis *transformer* yang dikombinasikan dengan strategi optimasi parameter dapat meningkatkan performa klasifikasi teks secara signifikan pada berbagai domain. Meskipun berbagai penelitian tersebut telah menunjukkan keberhasilan penggunaan model berbasis *transformer*, sebagian besar masih berfokus pada data media sosial atau domain umum seperti yang dilakukan oleh Maharani et al [9], Anderies et al [10], serta Kusoema et al [11], belum secara spesifik mengkaji kombinasi antara model IndoBERT dan *hyperparameter optimization* pada domain berita saham berbahasa Indonesia. Selain itu, performa model berbasis *deep learning* sangat dipengaruhi oleh konfigurasi *hyperparameter* yang digunakan, sehingga diperlukan pendekatan optimasi yang efektif [8].

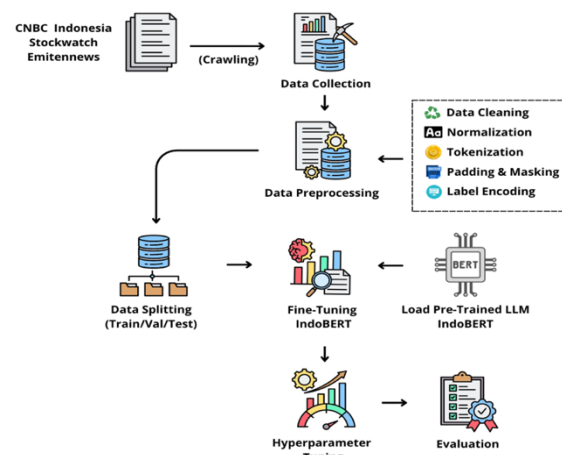
Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan model analisis sentimen berita saham berbahasa Indonesia menggunakan IndoBERT dengan pendekatan *fine-*

tuning dan *hyperparameter optimization*. Metode optimasi yang digunakan adalah *Bayesian Optimization* melalui *framework* Optuna untuk menemukan konfigurasi parameter terbaik. Dataset yang digunakan terdiri dari 23.108 artikel berita saham yang diklasifikasikan ke dalam tiga kelas sentimen, yaitu positif, netral, dan negatif. Kontribusi dalam penelitian ini yang pertama adalah penerapan *hyperparameter optimization* pada model IndoBERT untuk domain finansial berbahasa Indonesia, kemudian yang kedua adalah evaluasi komprehensif menggunakan berbagai metrik kinerja termasuk ROC-AUC dengan pendekatan *one-vs-rest*, serta yang ketiga yaitu identifikasi model terbaik untuk analisis sentimen berita saham yang dapat diimplementasikan dalam sistem monitoring pasar secara otomatis.

II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Design Science Research* (DSR) [12] yang berfokus pada pengembangan artefak berupa model analisis sentimen berbasis IndoBERT. Pendekatan ini mencakup tahapan identifikasi masalah, perancangan solusi, pengembangan model, serta evaluasi untuk memastikan performa yang optimal.

Untuk mengimplementasikan pendekatan tersebut secara teknis, penelitian ini dirancang dalam bentuk *pipeline* eksperimen yang terdiri dari beberapa tahapan utama, mulai dari pengumpulan data hingga evaluasi model. Alur implementasi penelitian ditunjukkan pada Gambar 1.



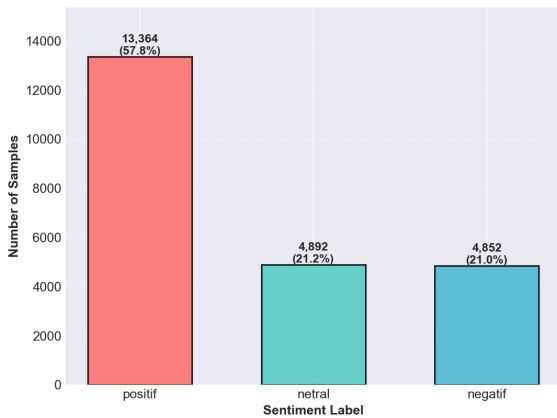
Gambar 1. Alur penelitian

A. Data Collection

Dataset yang digunakan terdiri dari 23.108 artikel berita saham berbahasa Indonesia yang dikumpulkan dari tiga portal berita finansial, yaitu CNBC Indonesia, Stockwatch.id, dan Emitennews.com.

Data dikumpulkan melalui proses web scraping pada periode Januari 2024 hingga Oktober 2025.

Setiap data terdiri dari judul dan isi berita yang kemudian dilabeli ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Proses pelabelan dilakukan secara manual oleh dua annotator untuk memastikan konsistensi label. Distribusi dataset ditunjukkan pada Gambar 2.



Gambar 2. Distribusi dataset

B. Preprocessing Data

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam pelatihan model. Proses ini difokuskan pada pembersihan teks dan normalisasi konten agar lebih representatif terhadap informasi utama dalam berita.

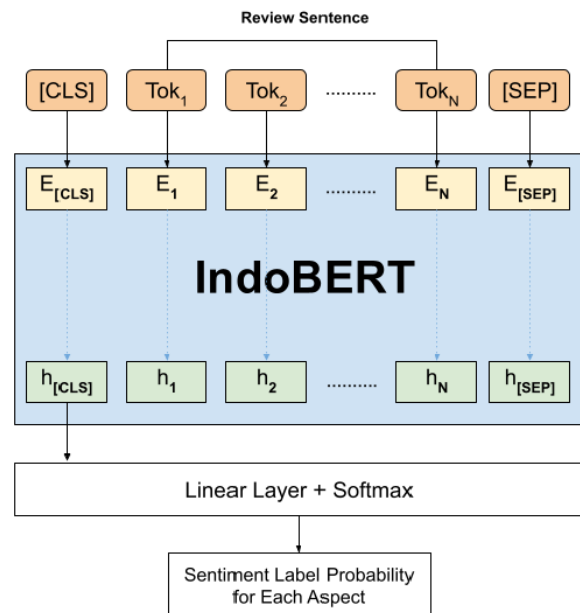
Proses *text cleaning* meliputi penghapusan karakter yang tidak relevan seperti URL, simbol, tanda baca berlebih, serta elemen *boilerplate* dari sumber berita, termasuk *prefix* nama media dan informasi lokasi. Selain itu, dilakukan normalisasi teks untuk menghilangkan *whitespace* yang tidak konsisten dan karakter tersembunyi [13]. Setelah proses pembersihan, teks diproses menggunakan *tokenizer* IndoBERT untuk mengubah teks menjadi representasi token yang sesuai dengan model. Proses ini juga mencakup *padding* dan *truncation* untuk menyeragamkan panjang *input* [14].

Dataset yang telah dipraproses kemudian dibagi menjadi tiga subset, yaitu *train set*, *validation set*, dan *test set*, dengan proporsi 70%, 15%, dan 15%. Pembagian dilakukan secara bertahap menggunakan metode *stratified split* [15] untuk menjaga proporsi kelas sentimen tetap seimbang pada setiap subset. Selain itu, ketidakseimbangan kelas pada data latih ditangani menggunakan *class weights* yang dihitung berdasarkan frekuensi masing-masing kelas dan diterapkan pada fungsi *loss*.

C. Model IndoBERT

Model yang digunakan dalam penelitian ini adalah IndoBERT sebagai *pre-trained language*

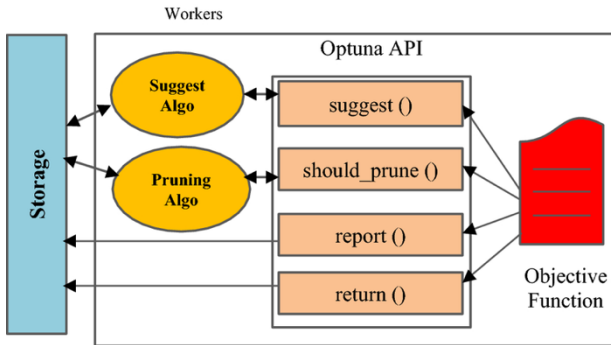
model berbasis arsitektur *transformer* yang telah dilatih pada korpus besar berbahasa Indonesia [7]. IndoBERT dipilih karena memiliki kemampuan representasi konteks yang baik dalam memahami struktur dan makna teks bahasa Indonesia, khususnya pada domain teks formal seperti berita. Proses pemodelan dilakukan dengan pendekatan *fine-tuning* [6], yaitu dengan menambahkan lapisan klasifikasi (*classification head*) pada token khusus [CLS] yang merepresentasikan keseluruhan input teks. Lapisan ini digunakan untuk menghasilkan prediksi tiga kelas sentimen, yaitu positif, netral, dan negatif. Alur proses *fine-tuning* model IndoBERT ditunjukkan pada Gambar 3.



Gambar 3. Arsitektur *Fine-Tuning* IndoBERT

D. Hyperparameter Optimization

Untuk meningkatkan performa model, dilakukan optimasi *hyperparameter* menggunakan pendekatan *Bayesian Optimization* [16] dengan memanfaatkan *framework* Optuna. Visualisasi detail dari *framework* optuna ini dapat dilihat pada Gambar 4. Pendekatan ini dipilih karena mampu melakukan pencarian parameter secara adaptif dan efisien berdasarkan hasil evaluasi pada iterasi sebelumnya [17].



Gambar 4. Desain arsitektur optuna

Dalam Optuna, proses optimasi dilakukan melalui serangkaian *trial* (percobaan) dimana setiap *trial* mencoba kombinasi *hyperparameter* yang berbeda. Model kemudian dievaluasi menggunakan fungsi objektif untuk menentukan performa dari konfigurasi parameter tersebut. Proses optimasi dapat direpresentasikan pada Persamaan 1.

$$\theta_{best} = \arg \min_{\theta \in \Theta} f(\theta) \quad (1)$$

Pada Persamaan 1, $f(\theta)$ merupakan fungsi objektif, dan θ_{best} *hyperparameter* terbaik. Proses optimasi dilakukan setelah tahap *fine-tuning* awal untuk menemukan kombinasi parameter terbaik yang dapat memaksimalkan kinerja model IndoBERT. Parameter yang dioptimasi meliputi *learning rate*, *batch size*, jumlah *epoch*, *weight decay*, *warmup ratio*, *gradient accumulation*, serta *dropout rate*.

Setiap kombinasi parameter diuji melalui beberapa percobaan (*trial*) dan dievaluasi menggunakan metrik *Macro-F1* pada data validasi. Penggunaan *Macro-F1* sebagai metrik utama dalam proses optimasi bertujuan untuk mengakomodasi ketidakseimbangan distribusi kelas pada dataset, sehingga performa model pada setiap kelas dapat dievaluasi secara lebih adil. Ruang pencarian *hyperparameter* yang digunakan dalam proses optimasi ditunjukkan pada Tabel 1.

Tabel 1. Parameter ruang pencarian

<i>Hyperparameter</i>	Rentang / Nilai
<i>Learning Rate</i>	$8 \times 10^{-6} - 1,5 \times 10^{-5}$
<i>Batch Size</i>	8 (<i>fixed</i>)
<i>Epoch</i>	8 – 10
<i>Dropout</i>	0.08 – 0.12
<i>Weight Decay</i>	0.008 – 0.015
<i>Warmup Ratio</i>	0.18 – 0.22
<i>Max Sequence Length</i>	256 (<i>fixed</i>)
<i>Gradient Accumulation</i>	{2, 4, 6}
<i>Label Smoothing</i>	0.0 (<i>fixed</i>)
<i>Focal Loss</i>	<i>Enabled</i>

<i>Alpha (Focal Loss)</i>	0.25
<i>Gamma (Focal Loss)</i>	2.0

E. Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa model dalam melakukan klasifikasi sentimen pada berita saham. Metrik evaluasi yang digunakan meliputi *accuracy*, *precision*, *recall*, dan *F1-score* [5]. *Accuracy* digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data, yang dirumuskan pada Persamaan 2.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Pada Persamaan 2, (TP) (*True Positive*) merupakan jumlah prediksi positif yang benar, (TN) (*True Negative*) merupakan jumlah prediksi negatif yang benar, (FP) (*False Positive*) merupakan jumlah prediksi positif yang salah, dan (FN) (*False Negative*) merupakan jumlah prediksi negatif yang salah. Selain *accuracy*, digunakan *precision* dan *recall* untuk mengukur ketepatan dan kelengkapan hasil klasifikasi. *Precision* menunjukkan proporsi prediksi positif yang benar, sedangkan *recall* menunjukkan kemampuan model dalam mendeteksi seluruh data positif. Kedua metrik tersebut dirumuskan pada Persamaan 3 dan Persamaan 4.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Kemudian metrik selanjutnya yang digunakan dalam evaluasi model adalah *F1-score* yang digunakan sebagai metrik utama karena mampu merepresentasikan keseimbangan antara *precision* dan *recall*. Untuk *F1-Score* ini dirumuskan pada Persamaan 5.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Berikutnya selain menggunakan metrik *F1-score*, untuk mengatasi ketidakseimbangan distribusi kelas, digunakan juga metrik evaluasi dengan *Macro-F1* yang dihitung sebagai rata-rata *F1-score* dari setiap kelas, yang dirumuskan sebagai Persamaan 6 dengan C merupakan jumlah kelas sentimen dan $F1_i$ adalah nilai *F1-score* pada kelas ke- i [18].

$$Macro F1 = \sum_{i=1}^c F1_i \quad (6)$$

Selanjutnya, evaluasi model juga dilakukan menggunakan *confusion matrix* [19] untuk menganalisis pola kesalahan klasifikasi pada setiap kelas. Untuk mengukur kemampuan model dalam membedakan antar kelas, digunakan kurva *Receiver Operating Characteristic* ROC dan nilai *Area Under the Curve* (AUC) dengan pendekatan *one-vs-rest* (OvR) pada kasus klasifikasi multi-kelas [20],[21]. Pendekatan OvR dilakukan dengan membandingkan setiap kelas terhadap kelas lainnya secara bergantian, sehingga diperoleh nilai AUC untuk masing-masing kelas yang kemudian dapat dianalisis secara komprehensif. Sebagai tambahan, untuk menangani ketidakseimbangan data pada proses pelatihan, digunakan *class weights* yang dihitung berdasarkan frekuensi masing-masing kelas pada data latih [22].

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksperimen serta pembahasan mengenai performa model dalam melakukan klasifikasi sentimen pada berita saham berbahasa Indonesia. Evaluasi dilakukan terhadap beberapa varian model IndoBERT dengan mempertimbangkan pengaruh penerapan *hyperparameter optimization* menggunakan Optuna.

Analisis difokuskan pada perbandingan performa model, konfigurasi parameter terbaik yang diperoleh dari proses optimasi, serta kemampuan model dalam

mengklasifikasikan dan membedakan antar kelas sentimen. Hasil evaluasi disajikan menggunakan berbagai metrik, antara lain *accuracy*, *precision*, *recall*, *F1-score*, dan *Macro-F1*, serta didukung oleh analisis *confusion matrix* dan ROC-AUC untuk memberikan gambaran yang lebih komprehensif terhadap performa model.

A. Konfigurasi Hyperparameter Terbaik

Proses *hyperparameter optimization* dilakukan menggunakan *framework* Optuna untuk menemukan kombinasi parameter terbaik yang dapat meningkatkan performa model IndoBERT. Hasil optimasi menunjukkan bahwa setiap model memiliki konfigurasi parameter optimal yang berbeda, tergantung pada karakteristik arsitektur dan data yang digunakan. Konfigurasi *hyperparameter* terbaik yang diperoleh dari proses optimasi ditunjukkan pada Tabel 2. Dari hasil tersebut, dapat dilihat bahwa pemilihan nilai *learning rate*, *batch size*, *epoch*, dan *weight decay* memiliki pengaruh signifikan terhadap performa model. Nilai *learning rate* yang relatif kecil membantu model dalam mencapai konvergensi yang stabil, sedangkan pengaturan *batch size* dan *epoch* berperan dalam menjaga keseimbangan antara kecepatan pelatihan dan kemampuan generalisasi model.

Selain itu, penerapan *weight decay* dan *dropout* juga berkontribusi dalam mengurangi risiko *overfitting* selama proses pelatihan. Hal ini menunjukkan bahwa kombinasi parameter yang optimal tidak hanya meningkatkan akurasi model, tetapi juga memperkuat stabilitas dan konsistensi performa pada data uji.

Tabel 2. Hasil konfigurasi *hyperparameter* terbaik menggunakan Optuna

No	Model	Learning Rate	Epoch	Grad. Acc.	Dropout	Weight Decay	Warmup	Focal Loss
1	IndoBERT Base Uncased	8.10×10^{-6}	9	2	0.1089	0.01457	0.1800	Ya
2	IndoBERT Base P2	8.10×10^{-6}	9	2	0.1089	0.01457	0.1800	Ya
3	IndoBERT Base Large P2	8.07×10^{-6}	8	2	0.1186	0.01366	0.1922	Ya
4	IndoBERT Base Lite P2	1.48×10^{-5}	8	6	0.0819	0.01225	0.1868	Ya

B. Perbandingan Performa Model

Eksperimen dilakukan untuk membandingkan performa model IndoBERT sebelum dan sesudah penerapan *hyperparameter optimization* menggunakan Optuna. Evaluasi dilakukan pada dataset uji dengan menggunakan metrik *accuracy*, *precision*, *recall*, dan *Macro-F1* untuk memberikan penilaian performa yang lebih komprehensif pada klasifikasi multi-kelas. Hasil perbandingan performa model ditunjukkan pada Tabel 3. Berdasarkan hasil tersebut, terlihat bahwa model yang telah dioptimasi

menunjukkan peningkatan performa dibandingkan model *baseline*. Peningkatan ini terlihat konsisten pada berbagai metrik evaluasi, terutama pada nilai *Macro-F1* yang lebih representatif dalam kondisi distribusi data yang tidak seimbang.

Hasil ini menunjukkan bahwa pemilihan konfigurasi *hyperparameter* yang optimal memiliki pengaruh signifikan terhadap kemampuan model dalam melakukan generalisasi pada data yang belum pernah dilihat sebelumnya.

Tabel 3. Perbandingan performa model

Model	Class	Precision	Recall	F1-score	Accuracy	Macro-F1	HPO
IndoBERT Base Uncased	Negatif	0.8381	0.8036	0.8205	0.8269	0.7816	Ya
	Netral	0.6463	0.6049	0.6249			
	Positif	0.8828	0.9167	0.8994			
IndoBERT Base P2	Negatif	0.8025	0.7871	0.7947	0.8030	0.7574	Ya
	Netral	0.5854	0.6022	0.5937			
	Positif	0.8854	0.8823	0.8838			
IndoBERT Large P2	Negatif	0.8192	0.7967	0.8078	0.8062	0.7625	Ya
	Netral	0.5970	0.5995	0.5982			
	Positif	0.8778	0.8853	0.8815			
IndoBERT Lite P2	Negatif	0.8266	0.7596	0.7917	0.7972	0.7501	Ya
	Netral	0.5627	0.5926	0.5773			
	Positif	0.8770	0.8858	0.8814			
IndoBERT Baseline	Negatif	0.7935	0.7992	0.7963	0.8057	0.7585	Tidak
	Netral	0.6019	0.5859	0.5938			
	Positif	0.8822	0.8885	0.8854			

C. Model Terbaik

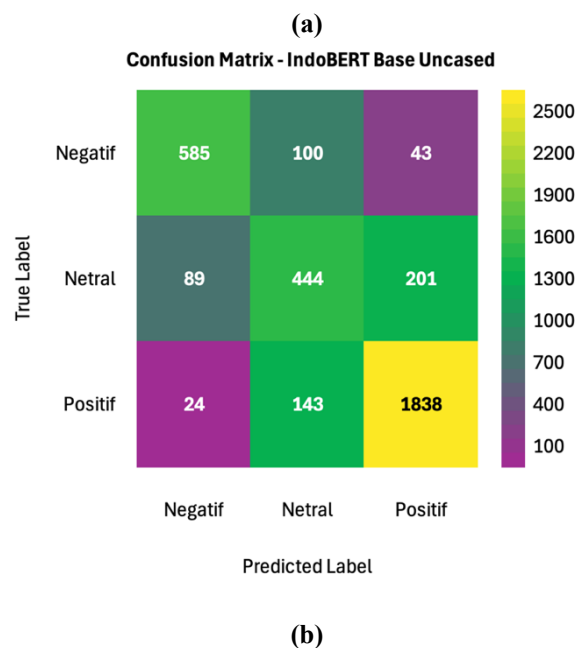
Berdasarkan hasil perbandingan performa model pada Tabel 3, model IndoBERT-Base-Uncased dengan konfigurasi *hyperparameter* optimal menunjukkan performa terbaik dibandingkan model lainnya. Model ini menghasilkan nilai *accuracy* sebesar 0,8269 dan *Macro-F1* sebesar 0,7816 pada dataset uji. Performa tersebut menunjukkan bahwa model mampu menangkap pola sentimen dalam teks berita saham dengan baik, meskipun terdapat variasi kompleks dalam struktur bahasa dan penggunaan istilah finansial. Peningkatan performa dibandingkan model baseline menunjukkan bahwa proses *hyperparameter optimization* berkontribusi signifikan dalam meningkatkan kualitas prediksi model. Selain itu, konfigurasi parameter yang optimal memungkinkan model untuk mencapai keseimbangan antara kemampuan generalisasi dan stabilitas selama proses pelatihan. Hal ini penting dalam konteks data berita saham yang memiliki karakteristik dinamis dan beragam. Dengan demikian, model IndoBERT-Base-Uncased yang telah dioptimasi dapat digunakan sebagai model utama dalam analisis sentimen berita saham berbahasa Indonesia.

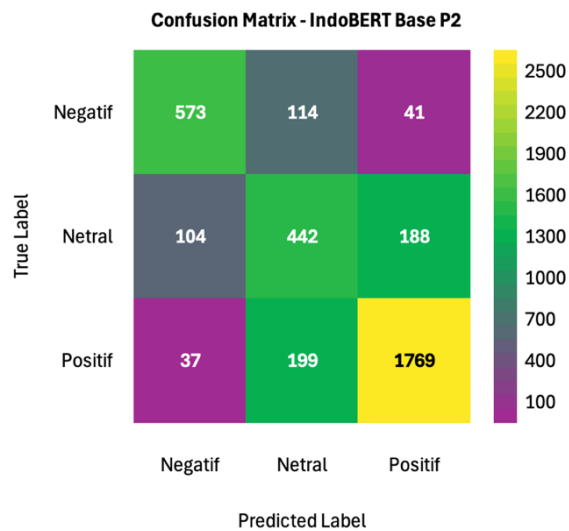
D. Analisis Confusion Matrix

Untuk menganalisis pola kesalahan klasifikasi model, digunakan *confusion matrix* seperti yang ditunjukkan pada Gambar 5. Pada Gambar 5 menampilkan *confusion matrix* dari beberapa varian model IndoBERT, yaitu (a) IndoBERT Base Uncased, (b) IndoBERT Base P2, (c) IndoBERT Large P2, dan (d) IndoBERT Lite P2.

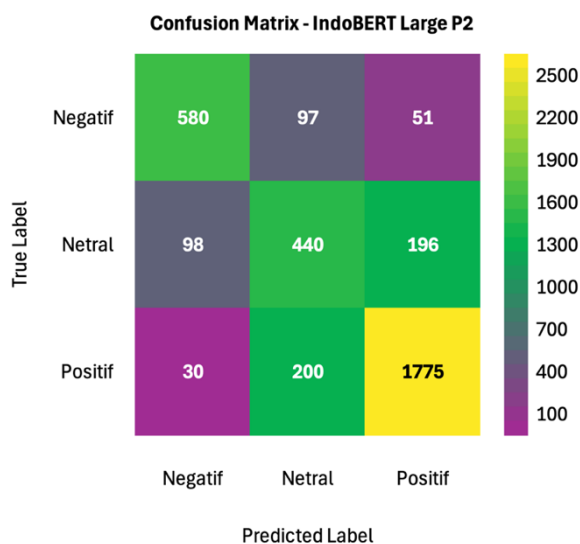
Secara umum, seluruh model menunjukkan performa yang baik dalam mengklasifikasikan kelas positif, yang ditunjukkan oleh nilai prediksi benar yang dominan pada diagonal utama. Selain itu, model juga mampu mempertahankan konsistensi

dalam mengklasifikasikan kelas negatif dengan tingkat kesalahan yang relatif rendah. Hal ini menunjukkan bahwa representasi kontekstual yang dihasilkan oleh model IndoBERT cukup efektif dalam memahami pola sentimen yang bersifat eksplisit. Pada kelas netral, terdapat variasi prediksi antar model yang menunjukkan adanya perbedaan sensitivitas model terhadap karakteristik teks yang bersifat informatif. Namun demikian, secara keseluruhan model tetap mampu mengklasifikasikan sebagian besar data dengan baik. Jika dibandingkan antar model, IndoBERT Base Uncased pada Gambar 5 (a) menunjukkan performa yang paling seimbang di antara ketiga kelas sentimen.

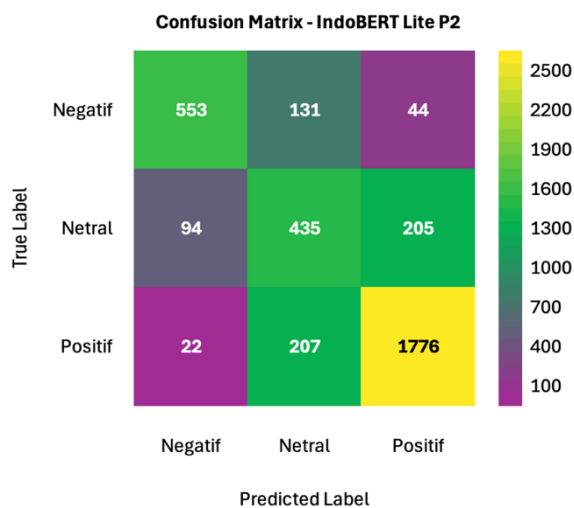




(c)



(d)



Gambar 5. Confusion Matrix dari setiap model, (a) IndoBERT Base Uncased, (b) IndoBERT Base P2, (c) IndoBERT Large P2, (d) IndoBERT Lite P2

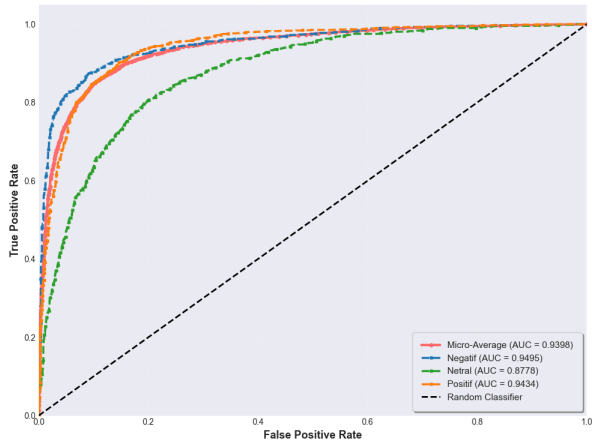
Sementara itu, varian model lainnya juga menunjukkan performa yang kompetitif dengan perbedaan yang tidak terlalu signifikan. Secara keseluruhan, hasil ini menunjukkan bahwa model berbasis IndoBERT memiliki kemampuan yang baik dalam melakukan klasifikasi sentimen pada berita saham, dengan performa yang relatif stabil di berbagai varian model.

E. Analisis ROC-AUC

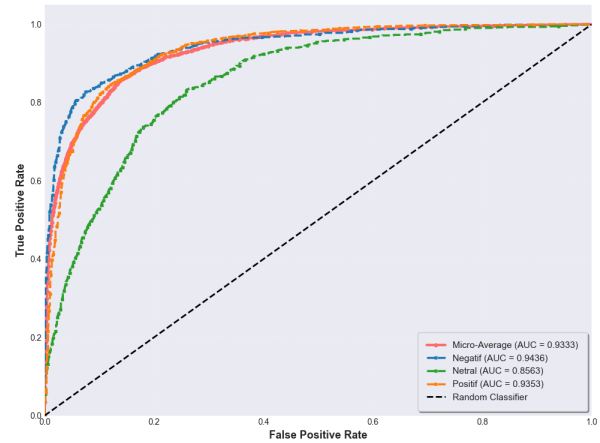
Selain menggunakan metrik berbasis *confusion matrix*, performa model juga dianalisis menggunakan kurva *Receiver Operating Characteristic* (ROC) dan nilai *Area Under the Curve* (AUC). Evaluasi ini bertujuan untuk mengukur kemampuan model dalam membedakan antar kelas sentimen secara lebih komprehensif. Pada kasus klasifikasi multi-kelas, ROC-AUC dihitung menggunakan pendekatan *one-vs-rest* (OvR), yaitu dengan membandingkan setiap kelas terhadap gabungan kelas lainnya secara bergantian. Hasil kurva ROC untuk masing-masing model ditunjukkan pada Gambar 6.

Berdasarkan hasil evaluasi, model IndoBERT Base Uncased Gambar 6 (a) menunjukkan performa terbaik dengan nilai *micro-average* AUC sebesar 0,9398. Secara lebih rinci, model ini menghasilkan nilai AUC sebesar 0,9485 untuk kelas negatif, 0,9739 untuk kelas netral, dan 0,9344 untuk kelas positif. Nilai ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antar kelas sentimen. Sementara itu, model IndoBERT Base P2 Gambar 6 (b) menunjukkan performa yang kompetitif dengan nilai *micro-average* AUC sebesar 0,9222, dengan rincian AUC sebesar 0,9579 untuk kelas negatif, 0,8512 untuk kelas netral, dan 0,9512 untuk kelas positif. Pada model IndoBERT Large P2 Gambar 6 (c), diperoleh nilai *micro-average* AUC sebesar 0,9375, dengan AUC masing-masing kelas sebesar 0,9543 untuk kelas negatif, 0,8609 untuk kelas netral, dan 0,9370 untuk kelas positif.

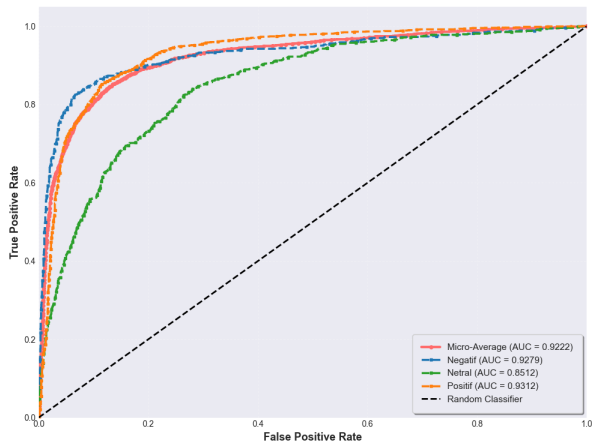
(a)



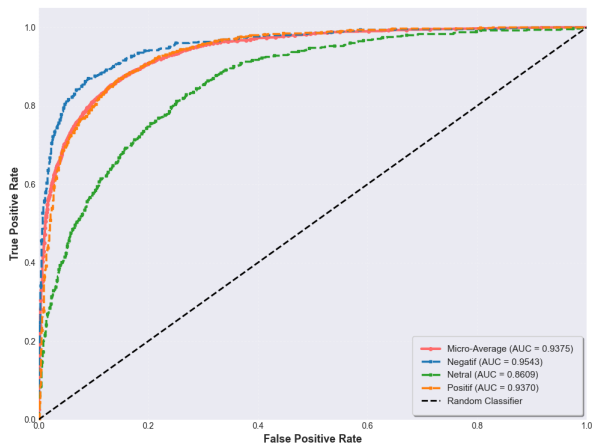
(b)



Gambar 6. ROC Curve & AUC dari setiap model, (a) IndoBERT Base Uncased, (b) IndoBERT Base P2, (c) IndoBERT Large P2, (d) IndoBERT Lite P2



(c)



(d)

Adapun model IndoBERT Lite P2 Gambar 6 (d) menghasilkan nilai *micro-average* AUC sebesar 0,9333, dengan rincian AUC sebesar 0,9436 untuk kelas negatif, 0,8563 untuk kelas netral, dan 0,9553 untuk kelas positif. Dari semua hasil tersebut, seluruh model menunjukkan nilai AUC yang tinggi (di atas 0,9), yang mengindikasikan kemampuan yang baik dalam memisahkan kelas sentimen. Jika dibandingkan antar model, IndoBERT Base Uncased menunjukkan performa yang paling konsisten, terutama dalam menjaga keseimbangan performa pada setiap kelas. Hasil ini memperkuat temuan sebelumnya bahwa model berbasis IndoBERT memiliki kemampuan yang baik dalam melakukan klasifikasi sentimen pada berita saham, dengan tingkat separasi kelas yang tinggi dan performa yang relatif stabil di berbagai varian model.

F. Pembahasan

Berdasarkan hasil eksperimen yang telah dilakukan, dapat disimpulkan bahwa penerapan *hyperparameter optimization* menggunakan Optuna memberikan peningkatan performa yang konsisten pada model IndoBERT. Hal ini terlihat dari peningkatan nilai *accuracy* dan *Macro-F1* dibandingkan dengan model baseline, serta diperkuat oleh hasil evaluasi ROC-AUC yang menunjukkan kemampuan separasi kelas yang tinggi.

Performa terbaik yang dicapai oleh model IndoBERT Base Uncased menunjukkan bahwa konfigurasi *hyperparameter* yang optimal berperan penting dalam meningkatkan kemampuan model dalam memahami konteks teks. Model berbasis *transformer* seperti IndoBERT memiliki keunggulan dalam menangkap dependensi kontekstual dalam teks, sehingga lebih efektif dalam mengolah informasi yang kompleks, seperti pada berita saham.

Selain itu, hasil analisis *confusion matrix* menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan baik, terutama pada kelas dengan pola sentimen yang lebih eksplisit. Sementara itu, variasi prediksi pada kelas tertentu menunjukkan bahwa model masih dipengaruhi oleh karakteristik data yang bersifat informatif dan kontekstual. Hasil evaluasi menggunakan ROC-AUC juga menunjukkan bahwa seluruh model memiliki kemampuan yang baik dalam membedakan antar kelas, dengan nilai AUC yang tinggi pada setiap kelas. Hal ini mengindikasikan bahwa model tidak hanya memiliki performa yang baik pada metrik berbasis klasifikasi, tetapi juga memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya.

Hasil akhir dari pembahasan, penelitian ini menunjukkan bahwa kombinasi antara model IndoBERT dan *hyperparameter optimization* merupakan pendekatan yang efektif untuk analisis sentimen pada berita saham berbahasa Indonesia. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa optimasi parameter dapat meningkatkan performa model berbasis *deep learning*, khususnya pada tugas klasifikasi teks. Meskipun demikian, hasil penelitian ini juga menunjukkan adanya peluang pengembangan lebih lanjut, seperti eksplorasi model berbasis domain khusus atau integrasi dengan metode lain untuk meningkatkan performa pada kondisi data yang lebih kompleks. Hal ini membuka peluang penelitian lanjutan dalam pengembangan sistem analisis sentimen yang lebih adaptif dan akurat.

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan model analisis sentimen pada berita saham berbahasa Indonesia menggunakan IndoBERT dengan pendekatan *hyperparameter optimization*. Hasil eksperimen menunjukkan bahwa optimasi menggunakan Optuna mampu meningkatkan performa model secara signifikan, dengan model terbaik IndoBERT Base Uncased mencapai *accuracy* sebesar 0,8269 dan *Macro-F1* sebesar 0,7816. Evaluasi tambahan menggunakan ROC-AUC juga menunjukkan kemampuan model yang baik dalam membedakan antar kelas sentimen. Secara keseluruhan, kombinasi model berbasis *transformer* dan optimasi *hyperparameter* terbukti efektif untuk analisis sentimen pada domain berita saham. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi model berbasis domain khusus serta pendekatan lanjutan guna

meningkatkan performa pada data dengan karakteristik yang lebih kompleks.

REFERENSI

- [1] KSEI, "Statistik Pasar Modal Indonesia – Agustus 2025," Jakarta, 2025. [Online]. Available: https://web.ksei.co.id/files/Statistik_Publik_Agustus_2025.pdf
- [2] A. Alamsyah *et al.*, "Deciphering news sentiment and stock price relationships in Indonesian companies: an AI-based exploration of industry affiliation and news co-occurrence," *Discov. Artif. Intell.*, vol. 5, no. 1, p. 87, Jun. 2025, doi: 10.1007/s44163-025-00350-5.
- [3] Y. Song, Y. Zhang, J. Huang, and A. Yang, "Volatility and value-at-risk forecasting using BERT and transformer models incorporating investors' textual sentiments," *Financ. Res. Lett.*, vol. 85, p. 108210, Nov. 2025, doi: 10.1016/j.frl.2025.108210.
- [4] Z. Liu, D. Huang, K. Huang, Z. Li, and J. Zhao, "FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, California: International Joint Conferences on Artificial Intelligence Organization, Jul. 2020, pp. 4513–4519. doi: 10.24963/ijcai.2020/622.
- [5] A. H. Chasanah and H. Al Azies, "Optimasi Hyperparameter Model Ensemble untuk Klasifikasi Sentimen Ulasan OVO," *JTERA (Jurnal Teknol. Rekayasa)*, vol. 10, no. 2, p. 95, Jan. 2026, doi: 10.31544/jtera.v10.i2.2025.95-104.
- [6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.
- [7] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, K.-F. Wong, K. Knight, and H. Wu, Eds., Suzhou, China: Association for Computational Linguistics, Dec. 2020, pp. 843–857. doi: 10.18653/v1/2020.aacl-main.85.
- [8] Anugerah Simanjuntak *et al.*, "Research and Analysis of IndoBERT Hyperparameter Tuning in Fake News Detection," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 13, no. 1, pp. 60–67, Feb. 2024, doi: 10.22146/jnteti.v13i1.8532.
- [9] N. P. I. Maharani, A. Purwarianti, Y. Yustiawan, and F. C. Rochim, "Domain-Specific Language

- Model Post-Training for Indonesian Financial NLP,” *Proc. Int. Conf. Electr. Eng. Informatics*, 2023, doi: 10.1109/ICEEI59426.2023.10346625.
- [10] Anderies, R. Rahutomo, and B. Pardamean, “Finetuning IndoBERT to Understand Indonesian Stock Trader Slang Language,” in *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, IEEE, Oct. 2021, pp. 42–46. doi: 10.1109/ICCSAI53272.2021.9609746.
- [11] W. J. Kusoema 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.
- [12] B. Kuechler and V. Vaishnavi, “On theory development in design science research: anatomy of a research project,” *Eur. J. Inf. Syst.*, vol. 17, no. 5, pp. 489–504, Oct. 2008, doi: 10.1057/ejis.2008.40.
- [13] M. Al-alshaqi, D. B. Rawat, and C. Liu, “A BERT-Based Multimodal Framework for Enhanced Fake News Detection Using Text and Image Data Fusion,” *Computers*, vol. 14, no. 6, p. 237, Jun. 2025, doi: 10.3390/computers14060237.
- [14] L. Afuan, N. Hidayat, H. Hamdani, H. Ismanto, B. C. Purnama, and D. I. Ramdhani, “Optimizing BERT Models with Fine-Tuning for Indonesian Twitter Sentiment Analysis,” *J. Wirel. Mob. Networks, Ubiquitous Comput. Dependable Appl.*, vol. 16, no. 2, pp. 248–267, Jun. 2025, doi: 10.58346/JOWUA.2025.I2.016.
- [15] J. Yao, A. Alabousi, and O. Mironov, “Evaluation of a BERT Natural Language Processing Model for Automating CT and MRI Triage and Protocol Selection,” *Can. Assoc. Radiol. J.*, vol. 76, no. 2, pp. 265–272, May 2025, doi: 10.1177/08465371241255895.
- [16] M. González-Duque, R. Michael, S. Bartels, Y. Zainchkovskyy, S. Hauberg, and W. Boomsma, “A survey and benchmark of high-dimensional Bayesian optimization of discrete sequences,” in *Advances in Neural Information Processing Systems*, A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, Eds., Curran Associates, Inc., 2024, pp. 140478–140508. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2024/file/fe0007fcfd707673660ec0f9014bc48e-Paper-Datasets_and_Benchmarks_Track.pdf
- [17] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY, USA: ACM, Jul. 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701.
- [18] S. D. Parameswari, M. Lubis, S. Suakanto, Y. Z. Ramadhan, R. N. Amanah, and R. A. Dila, “Studi Perbandingan Naïve Bayes dan Support Vector Machine (SVM) dalam Analisis Sentimen Pengguna Metaverse,” *J. Teknol. dan Manaj. Ind. Terap.*, vol. 4, no. 3, pp. 1059–1065, Sep. 2025, doi: 10.55826/jtmit.v4i3.1122.
- [19] S. Sathyanarayanan, “Confusion Matrix-Based Performance Evaluation Metrics,” *African J. Biomed. Res.*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.
- [20] A. S. Rizkia, W. Wufron, and F. F. Roji, “Analisis Sentimen Coretax: Perbandingan Pelabelan Data Manual, Transformers-Based, dan LEXICON-Based pada Performa IndoBERT,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, Jul. 2025, doi: 10.57152/malcom.v5i3.2151.
- [21] M. A. A. O. Putri, I. W. Sumarjaya, and I. G. N. L. Wijayakusuma, “Aspect-Based Sentiment Analysis of Reviews for Pandawa Beach Using Naive Bayes and SVM Methods,” *J. Appl. Informatics Comput.*, vol. 9, no. 2, pp. 305–313, Mar. 2025, doi: 10.30871/jaic.v9i2.9083.
- [22] M. Jefri, R. Fauzi, and R. Y. Fa’rifah, “Sentiment-Aware Feature Recommendations for Maternal Mental-Health Apps via IndoBERT and BERTopic on Indonesian TikTok Data,” in *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*, IEEE, Jun. 2025, pp. 575–580. doi: 10.1109/ICERA66156.2025.11087326.