

Analisis Performa Class Weight Dan Focal Loss Pada Model Indobert Untuk Klasifikasi Teks Depresi Berbahasa Indonesia

Rafi Jonathan Siger, Muhammad Naufal*, Farrikh Alzami

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ¹111202214760@mhs.dinus.ac.id, ²m.naufal@dsn.dinus.ac.id, ³alzami@dsn.dinus.ac.id

Email Penulis Korespondensi: m.naufal@dsn.dinus.ac.id

Submitted 25-02-2026; Accepted 31-03-2026; Published 28-04-2026

Abstrak

Perkembangan deteksi depresi dapat dilakukan menggunakan eksplorasi pada konten sosial media. Namun, klasifikasi teks indikasi depresi menghadapi tantangan utama berupa ketidakseimbangan distribusi kelas, yang dapat menurunkan kemampuan generalisasi model. Penelitian ini bertujuan untuk menganalisis bagaimana metode mengatasi ketidakseimbangan kelas terhadap kinerja model IndoBERT dalam klasifikasi teks indikasi depresi berbahasa Indonesia dengan menekankan analisis kestabilan pelatihan berdasarkan dinamika training loss dan validation loss. Dataset yang digunakan terdiri dari 3.863 data, data yang telah melalui proses pembersihan, penghapusan data duplikat, tokenisasi, encoding, dan pembagian data menjadi stratifikasi menjadi data latih, data validasi, dan data uji. Model IndoBERT-base-p1 dilakukan fine-tuning menggunakan tiga skenario pelatihan, yaitu baseline, class weight, dan focal loss dengan mekanisme early stopping berbasis validation loss. Hasil pengujian menunjukkan bahwa skenario IndoBERT baseline menghasilkan akurasi sebesar 77,52%, precision berbobot sebesar 0,7752, recall berbobot sebesar 0,7752, F1-score berbobot sebesar 0,7737, serta ROC-AUC sebesar 0,8528 dengan pola pelatihan yang relatif stabil. Metode class weight menghasilkan akurasi sebesar 74,68%, F1-score berbobot sebesar 0,7467, dan ROC-AUC sebesar 0,8342 yang menunjukkan peningkatan kemampuan diskriminasi kelas namun disertai menurunkan akurasi keseluruhan. Sementara itu, metode focal loss menghasilkan akurasi sebesar 72,87%, F1-score berbobot sebesar 0,7291, serta ROC-AUC sebesar 0,8188 dengan karakteristik pelatihan yang lebih seimbang dibandingkan class weight. Temuan penelitian menunjukkan bahwa penanganan ketidakseimbangan kelas tidak selalu meningkatkan performa global, sehingga evaluasi model perlu mempertimbangkan keseimbangan antara akurasi, sensitivitas, dan stabilitas pelatihan.

Kata Kunci: Klasifikasi Teks; Indikasi Depresi; IndoBERT; Class Weight; Focal Loss

Abstract

The development of depression detection can be done using exploration of social media content. However, the classification of depression indicative texts faces a major challenge in the form of class distribution imbalances, which can degrade the model's generalization capabilities. This study aims to analyze how the method of overcoming class imbalance affects the performance of the IndoBERT model in the classification of Indonesian depression indication texts by emphasizing the analysis of training stability based on the dynamics of training loss and validation loss. The dataset used consists of 3,863 data, data that has gone through the process of cleaning, removing duplicate data, tokenization, encoding, and dividing data into stratification into training data, validation data, and test data. The IndoBERT-base-p1 model was fine-tuned using three training scenarios, namely baseline, class weight, and focal loss with an early stopping mechanism based on validation loss. The test results showed that the baseline IndoBERT scenario produced an accuracy of 77.52%, a weighted precision of 0.7752, a weighted recall of 0.7752, a weighted F1-score of 0.7737, and a ROC-AUC of 0.8528 with a relatively stable training pattern. The class weight method produced an accuracy of 74.68%, a weighted F1-score of 0.7467, and a ROC-AUC of 0.8342 which showed an increase in class discrimination ability but accompanied by a decrease in overall accuracy. Meanwhile, the focal loss method produced an accuracy of 72.87%, a weighted F1-score of 0.7291, and a ROC-AUC of 0.8188 with more balanced training characteristics than the weight class. The findings suggest that handling classroom imbalances does not necessarily improve global performance, so model evaluations need to consider a balance between accuracy, sensitivity, and stability of training.

Keywords: Text Classification; Depression Indication; IndoBERT; Class Weight; Focal Loss

1. PENDAHULUAN

Salah satu pendekatan penting dalam bidang pemrosesan bahasa alami (*Natural Language Processing/NLP*) adalah analisis teks indikasi depresi, yang bertujuan untuk menemukan pola bahasa yang menunjukkan gejala depresi dalam data teks berbahasa Indonesia [1]. Perkembangan platform media sosial seperti *Twitter (X)*, *Instagram*, dan *TikTok* menghasilkan volume data teks yang sangat besar di era teknologi saat ini. Pada tahun 2025, diproyeksikan lebih dari 212 juta pengguna internet dan 143 juta pengguna media sosial di Indonesia, sehingga membuka peluang besar pemanfaatan data teks sebagai sumber informasi untuk deteksi dini kesehatan mental [2]. Namun, besarnya volume data tersebut juga menimbulkan tantangan tersendiri dalam hal pengolahan, kualitas data, dan akurasi model klasifikasi.

Berbagai pendekatan telah dikembangkan untuk klasifikasi teks depresi, termasuk metode pengajaran mesin konvensional seperti *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *random Forest*, metode pengajaran mendalam seperti *Convolutional Neural Network (CNN)*, *Long Short-Term Memory (LSTM)*, dan *Gated Recurrent Unit (GRU)*, dan model transformer kontemporer seperti *RoBERTa*, *ALBERT*, *DistilBERT*, *XLM-RoBERTa*, dan *mBERT*[3]. Salah satu model yang dirancang khusus untuk bahasa Indonesia adalah *IndoBERT (Indonesian BERT)*, yaitu varian *BERT* yang dilatih menggunakan korpus berskala besar dengan pendekatan *dynamic masking* dan disesuaikan dengan fitur bahasa Indonesia seperti slang, singkatan dan gabungan kode yang sering terjadi di media sosial [4]. *IndoBERT* lebih baik dalam memahami konteks semantik bahasa Indonesia daripada model multibahasa yang umum seperti *Multilingual BERT (mBERT)*, *XLM-RoBERTa* dan *Multilingual DistilBERT*.

Meskipun demikian, permasalahan utama dalam klasifikasi teks indikasi depresi adalah ketidakseimbangan distribusi kelas (*class imbalance*), di mana jumlah teks yang tidak terindikasi depresi jauh lebih dominan dibandingkan teks yang terindikasi depresi, dengan rasio yang dapat mencapai 90:10 [5]. Ketika ini terjadi, model cenderung bias terhadap kelas mayoritas dan memiliki nilai akurasi yang tinggi secara keseluruhan, tetapi kurang efektif untuk mengidentifikasi kelas minoritas, yang lebih penting untuk kesehatan mental. Untuk mengatasi permasalahan tersebut, berbagai metode penanganan imbalance telah dikembangkan untuk menyeimbangkan data, seperti *oversampling* (*Random Oversampling*, *SMOTE*, *ADASYN*), *undersampling* (*Random Undersampling*, *Tomek Links*), *sampling* campuran, augmentasi data berbasis *NLP*, pembelajaran dengan sensitivitas biaya, *ensemble learning* berbasis penyeimbangan, penimbangan kelas, serta modifikasi fungsi kerugian seperti *focal loss* [6].

Dalam penelitian ini, metode *class weight* dan *focal loss* dipilih karena keduanya dapat diterapkan secara langsung pada proses pelatihan model *transformer* tanpa mengubah distribusi data asli. Pendekatan berbasis *sampling* berpotensi mengubah representasi semantik teks atau menyebabkan duplikasi data yang dapat meningkatkan risiko *overfitting*, terutama pada dataset dengan jumlah data terbatas. *Class weight* digunakan untuk meningkatkan kontribusi kesalahan pada kelas minoritas selama proses optimisasi, *focal loss* dirancang agar pembelajaran lebih fokus pada sampel yang sulit diklasifikasikan, sehingga model menjadi lebih sensitif terhadap kelas minoritas.

Sejumlah penelitian terdahulu menunjukkan bahwa *fine-tuning IndoBERT* dapat memberikan performa yang baik dalam tugas deteksi depresi [2], [7]. Namun, sebagian besar penelitian tersebut masih menitikberatkan pada pencapaian nilai akurasi akhir tanpa melakukan analisis mendalam terhadap kestabilan pelatihan model dan dinamika *loss* pada kondisi dataset yang tidak seimbang. Selain itu, beberapa penelitian juga mengasumsikan bahwa dataset telah bersih dan bebas dari permasalahan kualitas data, padahal pada data media sosial sering ditemukan data duplikat dan potensi data leakage yang dapat menyebabkan hasil evaluasi menjadi bias [8]. Evaluasi berbasis akurasi saja juga dinilai kurang representatif pada kondisi *class imbalance* karena tidak sepenuhnya menunjukkan kemampuan model dalam mengenali kelas yang jumlahnya sedikit.

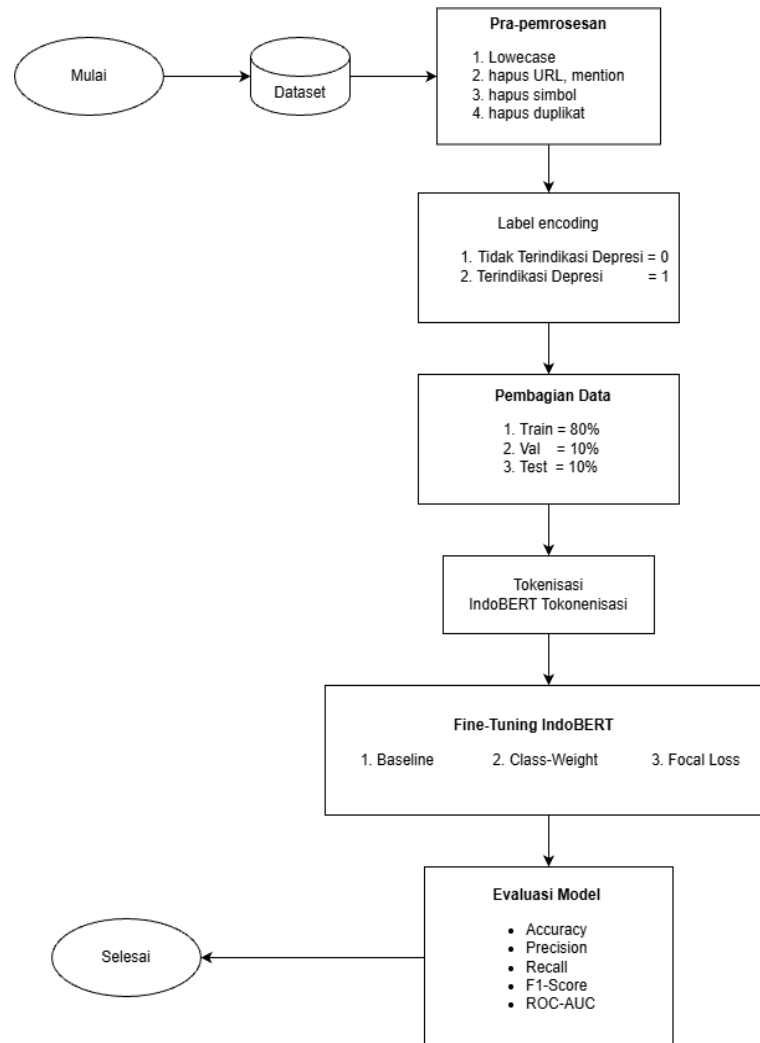
Penelitian ini menggunakan dataset teks yang berisi indikasi depresi dalam bahasa Indonesia, yang didapatkan dari penelitian sebelumnya, dataset tersebut terdiri dari dua kelas, yaitu Tidak Terindikasi Depresi dan Terindikasi Depresi, dengan jumlah kelas yang tidak seimbang [9]. Berbeda dengan penelitian sebelumnya, penelitian ini secara *eksplisit* menekankan pentingnya analisis kualitas data, penanganan ketidakseimbangan kelas, serta evaluasi berbasis *loss function* dan *metrik kurva* pada proses *fine-tuning IndoBERT*. Sebab itu, penelitian ini bertujuan untuk mengevaluasi dampak metode penanganan ketidakseimbangan kelas terhadap kinerja model *IndoBERT* dalam mengklasifikasikan teks indikasi depresi berbahasa Indonesia. Kontribusi utama penelitian ini adalah memberikan evaluasi secara lebih akurat kinerja model pada data yang tidak seimbang, serta membuktikan bahwa penerapan *focal loss* mampu meningkatkan stabilitas pelatihan dan performa model dalam mengenali teks terindikasi depresi. Hasil penelitian ini diharapkan dapat menjadi referensi dalam membuat sistem deteksi dini kesehatan mental berbasis teks yang lebih adil, peka, dan bisa diandalkan di Indonesia.

2. METODOLOGI PENELITIAN

Penelitian ini adalah penelitian berupa eksperimen kuantitatif yang bertujuan untuk menganalisis dampak penggunaan metode penanganan ketidakseimbangan kelas terhadap kinerja model *IndoBERT* dalam tugas klasifikasi teks indikasi depresi berbahasa Indonesia [10]. Pendekatan eksperimen dilakukan dengan membandingkan tiga skenario pelatihan model, yaitu *IndoBERT* baseline tanpa penanganan ketidakseimbangan kelas, *IndoBERT* dengan penerapan *class weight*, serta *IndoBERT* dengan penerapan *focal loss*.

Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada pencapaian nilai akurasi yang tinggi, penelitian ini menitikberatkan analisis pada kestabilan proses pelatihan model serta kemampuan model dalam mengenali kelas minoritas ketika dataset tidak seimbang. Oleh karena itu, pengevaluasian kemampuan model tidak hanya menggunakan metrik evaluasi biasa, tetapi juga melalui pengamatan terhadap nilai *loss* pada fase pelatihan dan validasi sebagai tanda apakah model stabil dan dapat berkonvergen selama proses *fine-tuning*.

Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1. Tahapan penelitian dimulai dari pengumpulan dataset, analisis awal untuk mengidentifikasi ketidakseimbangan kelas dan keberadaan data duplikat, pra-pemrosesan teks, *tokenisasi* dan *encoding*, pembagian dataset, proses *fine-tuning* model *IndoBERT* dengan tiga skenario pelatihan, hingga tahap evaluasi kinerja model.



Gambar 1. Alur Diagram Penelitian

2.1 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini adalah kumpulan teks indikasi depresi dalam bahasa Indonesia, yang didapatkan dari penelitian sebelumnya. Dataset awal terdiri dari 10.000 data teks yang berasal dari media sosial dan dibagi ke dalam kategori, yaitu Tidak terindikasi Depresi dan Terindikasi Depresi [9]. Distribusi kelas dalam dataset tidak seimbang, yaitu kelas Tidak Terindikasi Depresi memiliki jumlah data yang lebih banyak dibandingkan kelas Terindikasi Depresi.

Analisis awal terhadap dataset menunjukkan adanya data duplikat dengan konten teks dan label yang identik. Keberadaan data duplikat berpotensi menyebabkan bias pelatihan model serta menghasilkan estimasi performa yang tidak realistis [11]. Selain itu, apabila proses pembersihan data dilakukan setelah pembagian dataset, terdapat risiko terjadinya kebocoran data (*data leakage*) antara data latih dan data uji [12].

Sebab itu, seluruh proses pembersihan dataset, termasuk penghapusan data duplikat, dilakukan sebelum tahap pembagian data. Contoh dataset awal sebelum dilakukan pembersihan ditunjukkan pada Tabel 1.

Tabel 1. Dataset Awal

No	Teks Asli	Label
1	Aku capek hidup, rasanya kosong terus 😞	Terindikasi Depresi
2	Aku capek hidup, rasanya kosong terus 😞	Terindikasi Depresi
3	Hidup itu harus disyukuri walaupun capek	Tidak Terindikasi Depresi
4	Kadang pengen nyerah aja sama semuanya	Terindikasi Depresi
5	Alhamdulillah hari ini bahagia banget 😊	Tidak Terindikasi Depresi

2.2 Prapemrosesan Data

Proses pra-pemrosesan data dilakukan agar kualitas teks meningkat dan mengurangi gangguan yang bisa mengurangi kinerja model [13]. Proses pra-pemrosesan meliputi beberapa tahapan, yaitu:

1. Konversi seluruh teks ke huruf kecil (*lowercase*)

2. Penghapusan URL, mention, hashtag, emoji, serta simbol non-alfabet
3. Penghapusan spasi berlebih
4. Penghapusan data duplikat berdasarkan kesamaan konten teks

Sebelum pembagian data latih, validasi dan uji dilakukan, data duplikat dihapus. Ini dilakukan untuk mencegah data *leakage*, yang dapat menyebabkan model mengevaluasi data yang secara tidak langsung telah dipelajari selama proses pelatihan. Dataset setelah melalui tahapan pra-pemrosesan ditunjukkan pada Tabel 2.

Tabel 2. Dataset Setelah Pra Pemrosesan

No	Teks Setelah Preprocessing	Label
1	aku capek hidup rasanya kosong terus	Terindikasi Depresi
2	hidup itu harus disyukuri walaupun capek	Tidak Terindikasi Depresi
3	kadang pengen nyerah aja sama semuanya	Terindikasi Depresi
4	alhamdulillah hari ini bahagia banget	Tidak Terindikasi Depresi

Setelah proses pembersihan data duplikat, diperoleh sebanyak 3.863 data teks yang digunakan sebagai dataset akhir. Dataset tersebut tetap menunjukkan kondisi ketidakseimbangan kelas, di mana kelas Tidak Terindikasi Depresi berjumlah 2.136 data, sedangkan kelas Terindikasi Depresi berjumlah 1.727 data. Perbedaan distribusi kelas ini menunjukkan dominasi kelas mayoritas, sehingga dapat menyebabkan bias model selama pelatihan jika tidak ditangani dengan benar. Kondisi tersebut menjadi dasar penerapan metode penanganan ketidakseimbangan kelas pada tahap pelatihan model. Pelabelan data dilakukan dengan melakukan encoding label ke dalam bentuk numerik. Kelas Terindikasi diberi label 1, sedangkan kelas Tidak Terindikasi Depresi diberi label 0 [14]. Proses pelabelan ini bertujuan untuk menyesuaikan format data dengan kebutuhan model klasifikasi biner serta memudahkan proses komputasi selama tahap pelatihan dan evaluasi model [15].

2.3 Pembagian Data

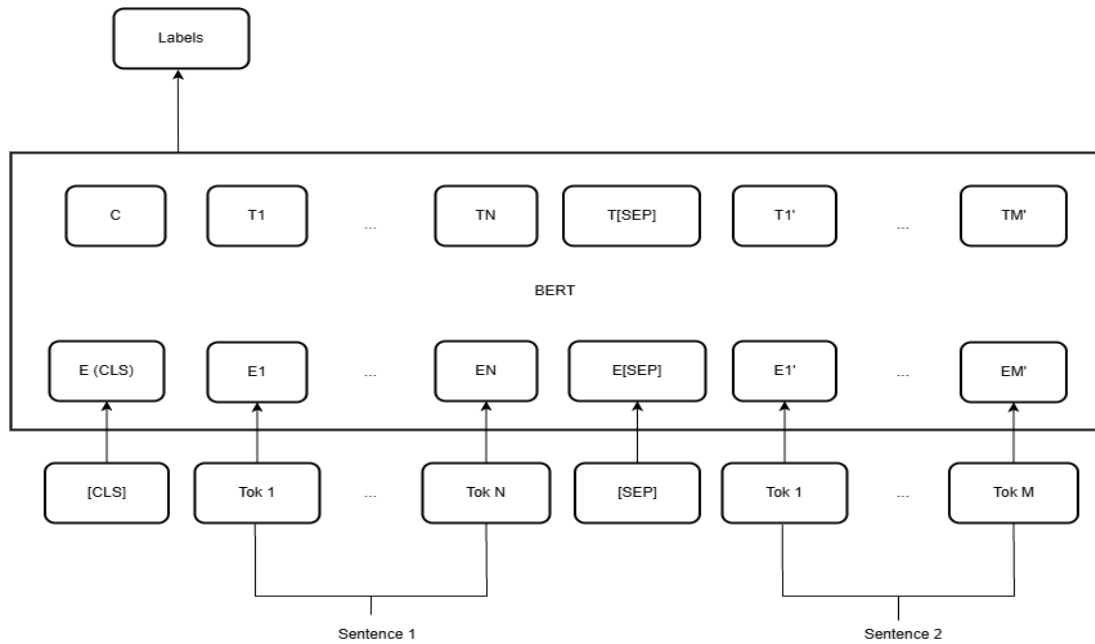
Untuk mempertahankan keseimbangan kelas, data yang sudah diproses dan diberi label dibagi menjadi tiga bagian kecil menggunakan metode *stratified* [16]. Data dibagi menjadi tiga bagian, yaitu data latih 80%, data validasi 10%, dan data uji 10%. Data latih digunakan untuk memperbaiki model secara lebih baik, data validasi digunakan untuk mengawasi proses pelatihan dan mencegah model terlalu mengingat data tertentu, sedangkan data uji digunakan untuk menilai hasil akhir model secara adil dan objektif. Pembagian data dilakukan satu kali dan digunakan secara konsisten pada seluruh skenario eksperimen.

2.4 Tokenisasi dan Encoding

Teks yang telah dibersihkan kemudian di tokenisasi menggunakan *tokenizer IndoBERT*. Proses tokenisasi melibatkan pemecahan teks menjadi token, penambahan token khusus seperti *[CLS]* dan *[SEP]*, serta penerapan *padding* dan *truncation* agar seluruh data memiliki panjang urutan yang sama [17]. Panjang maksimum urutan token ditetapkan sebesar 128 token. Hasil tokenisasi berupa *input IDs* dan *attention mask* yang digunakan sebagai masukan ke dalam model *IndoBERT* [18].

2.5 Arsitektur Model IndoBERT

Model yang digunakan dalam penelitian ini adalah *IndoBERT-base-p1*, yaitu model berbasis *transformer* yang merupakan versi yang sudah dilatih sebelumnya menggunakan kumpulan data bahasa Indonesia yang sangat besar [19]. Model ini mampu memahami arti dan konteks kalimat dalam bahasa Indonesia secara baik, termasuk bahasa informal, singkatan, dan ekspresi emosional yang umum ditemukan pada media sosial [20]. *IndoBERT* digunakan sebagai *backbone* model, kemudian ditambahkan sebuah lapisan klasifikasi linear pada bagian akhir untuk melakukan tugas klasifikasi biner. Struktur arsitektur *IndoBERT* mengikuti mekanisme dasar *BERT* yang terdiri dari *embedding layer* (*token embedding*, *position embedding*, dan *segment embedding*) serta beberapa lapisan *transformer encoder* yang memanfaatkan *multi-head self-attention* dan *feed-forward network* untuk menghasilkan representasi kontekstual teks. Kalimat diperoleh dari token khusus *[CLS]* yang ditempatkan pada awal input dan digunakan sebagai fitur utama pada proses klasifikasi [21]. Arsitektur model yang digunakan dalam penelitian ini, yaitu *IndoBERT* untuk klasifikasi teks yang ditunjukkan pada Gambar 2.



Gambar 2. Arsitektur IndoBERT

Pada skenario *baseline*, model *IndoBERT-base-p1* dilatih menggunakan fungsi kerugian *Cross-Entropy* standar tanpa penyesuaian terhadap distribusi kelas. Skenario ini berfungsi sebagai acuan awal untuk mengevaluasi dampak penerapan metode penanganan ketidakseimbangan kelas pada skenario pelatihan berikutnya. Proses pelatihan model menggunakan *optimizer AdamW*, yang merupakan pengembangan dari *Adam* dengan penambahan *weight decay* untuk meningkatkan generalisasi model [22]. Secara matematis, pembaruan parameter *AdamW* dirumuskan sebagai berikut:

$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} - \eta \lambda \theta_{t-1} \quad (1)$$

Pembagian dataset pada data latih, data validasi, data data uji dilakukan menggunakan metode *stratified split* untuk menjaga proporsi distribusi kelas tetap seimbang pada setiap subset data. Hasil pembagian dataset tersebut ditunjukkan pada tabel 3.

Tabel 3. Parameter Pelatihan

Parameter	Nilai
Optimizer	AdamW
Learning rate	2×10^{-5}
Batch size	32
Epoch	25
Weight decay	0.01
Max sequence length	128

Berdasarkan Tabel 3, sebagian besar data digunakan sebagai data latih untuk memaksimalkan proses pembelajaran model, sementara data validasi dan data uji digunakan untuk mengevaluasi kinerja model secara objektif dan menghindari bias selama proses pelatihan.

2.6 Early Stopping

Selain pengaturan parameter pelatihan, penelitian ini juga menerapkan mekanisme *early stopping* untuk mencegah terjadinya *overfitting* selama proses *fine-tuning* model *IndoBERT*. *Early stopping* digunakan untuk menghentikan proses pelatihan secara otomatis ketika hasil model pada data validasi tidak meningkat selama beberapa *epoch* beruntun [23]. Pada penelitian ini, evaluasi model dilakukan pada setiap *epoch* dengan memantau nilai *validation loss (eval_loss)* sebagai indikator utama kestabilan pelatihan dan kemampuan generalisasi model. *Mekanisme early stopping* diterapkan menggunakan nilai *patience* sebesar 3 *epoch*, yang berarti proses pelatihan akan dihentikan apabila *validation loss* tidak mengalami penurunan selama tiga *epoch* berturut-turut.

Selain itu, konfigurasi pelatihan menggunakan opsi *load_best_model_at_end=True*, sehingga model dengan nilai *validation loss* terbaik secara otomatis dipilih sebagai model akhir. Pendekatan ini bertujuan untuk memastikan bahwa model yang digunakan untuk evaluasi bukanlah model pada *epoch* terakhir, melainkan model dengan performa generalisasi terbaik berdasarkan data validasi.

Penerapan *early stopping* dalam penelitian ini berfungsi untuk mengurangi risiko *overfitting*, menjaga kestabilan proses pelatihan pada kondisi dataset yang tidak seimbang, serta meningkatkan efisiensi komputasi dengan menghentikan pelatihan ketika model tidak lagi menunjukkan peningkatan performa yang signifikan [24].

2.7 Class Weight

Fungsi *Cross-Entropy Loss* dimodifikasi dengan pemberian bobot kelas (*class weight*) yang dihitung berdasarkan distribusi kelas pada data latih [25]. Pemberian bobot kelas bertujuan untuk mengurangi dominasi kelas mayoritas serta meningkatkan kontribusi kesalahan dari kelas minoritas selama proses pelatihan. Secara matematis, fungsi kerugian dengan *class weight* dirumuskan sebagai berikut:

$$L_{CE} = - \sum_{i=1}^C w_i y_i \log \log (\hat{y}_i) \quad (2)$$

2.8 Focal Loss

Skenario ketiga menggunakan *Focal Loss* sebagai metode untuk memfokuskan proses pembelajaran pada sampel yang sulit diklasifikasikan [26]. Pendekatan ini dirancang untuk menurunkan kontribusi sampel yang mudah diklasifikasikan serta meningkatkan sensitivitas model terhadap kelas minoritas pada kondisi dataset yang tidak seimbang. Fungsi *Focal Loss* dirumuskan sebagai berikut:

$$L_{FL} = -\alpha(1 - p_t)^\gamma \log \log (p_t) \quad (3)$$

2.9 Evaluasi Model

Evaluasi kemampuan model dalam penelitian ini menggunakan beberapa ukuran penilaian untuk klasifikasi biner, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Beberapa metrik digunakan agar bisa menilai seberapa baik model bekerja secara lengkap, terutama ketika data yang digunakan tidak seimbang antara kelas Tidak Terindikasi Depresi dan Terindikasi Depresi. Selain itu, analisis *training loss* dan *validation loss* digunakan untuk mengevaluasi kestabilan proses pelatihan serta kemampuan generalisasi model selama proses *fine-tuning* berlangsung.

Evaluasi model menggunakan confusion matrix yang terdiri dari *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. *True Positive* adalah jumlah data yang masuk ke kategori Depresi dan berhasil diprediksi sebagai depresi. *True Negative* adalah jumlah data yang tidak masuk ke kategori depresi dan berhasil diprediksi sebagai tidak depresi. *False Positive* adalah data yang sebenarnya tidak depresi, tetapi salah prediksi sebagai depresi. sementara, *False Negative* adalah data yang sebenarnya depresi, tetapi salah diprediksi sebagai tidak depresi. *Accuracy* digunakan untuk mengukur berapa besar jumlah prediksi yang benar dibandingkan dengan seluruh data yang diuji. Rumus *accuracy* sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision digunakan untuk mengetahui seberapa tepat model dalam memprediksi kelas yang positif, yaitu kelas Terindikasi Depresi. *Precision* menjadi penting untuk mengetahui seberapa banyak prediksi depresi yang benar-benar sesuai dengan kondisi sebenarnya. Rumus *precision* adalah sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall digunakan untuk mengetahui sejauh mana kemampuan model dalam mendeteksi semua data yang sebenarnya masuk ke dalam kelas Depresi Terindikasi. Matrik ini sangat penting dalam konteks deteksi dini kesehatan mental karena berkaitan dengan kemampuan model untuk mengenali kasus depresi dengan sebaik mungkin [27]. Rumus *recall* dinyatakan sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

F1-score adalah rata-rata harmonik dari *precision* dan *recall* yang digunakan untuk mengukur seimbangannya ketepatan model dalam memprediksi dan kemampuannya dalam mendeteksi kelas positif [28]. Metrik ini sangat relevan digunakan pada kondisi dataset tidak seimbang. Rumus *F1-score* adalah sebagai berikut:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

Selain metrik evaluasi berbasis *confusion matrix*, penelitian ini juga menggunakan analisis *training loss* dan *validation loss* pada setiap *epoch* untuk mengevaluasi kestabilan proses pelatihan model. *Training loss* digunakan untuk mengetahui seberapa besar kesalahan yang terjadi saat model dilatih dengan data latih, sementara *validation loss* digunakan untuk melihat seberapa baik model bisa diterapkan pada data yang belum pernah digunakan selama proses pelatihan. Pola perubahan kedua nilai loss tersebut dianalisis untuk mengidentifikasi indikasi *overfitting*, *underfitting*, serta konvergensi model selama proses *fine-tuning* IndoBERT.

Penggunaan *confusion matrix* dalam penelitian ini bertujuan untuk menunjukkan secara jelas bagaimana kesalahan klasifikasi terdistribusi pada setiap kelas, sehingga evaluasi performa model, khususnya pada kelas minoritas dalam

kondisi dataset tidak seimbang, dapat dilakukan secara lebih komprehensif dan tidak hanya bergantung pada nilai akurasi secara keseluruhan.

3. HASIL DAN PEMBAHASAN

Bab ini menampilkan hasil dari penelitian serta pembahasan kinerja model *IndoBERT* dalam klasifikasi teks berbahasa Indonesia yang mengandung indikasi depresi. Sesuai dengan metodologi penelitian pada pembahasan sebelumnya, penilaian kinerja model tidak hanya didasarkan pada matrik klasifikasi akhir, tetapi juga mempertimbangkan analisis dari *training loss* dan *validation loss* sebagai indikator utama kestabilan proses pelatihan dan kemampuan generalisasi model pada kondisi dataset yang tidak seimbang.

Eksperimen dilakukan melalui beberapa tahapan utama, yaitu preprocessing data teks untuk membersihkan dan menormalisasi data, pembagian dataset menggunakan metode *stratified split*, serta proses *fine-tuning* model *IndoBERT* menggunakan tiga skenario pelatihan, yaitu *IndoBERT baseline* tanpa penanganan ketidakseimbangan kelas, *IndoBERT* dengan penerapan *class weight*, dan *IndoBERT* dengan penerapan *focal loss*. Seluruh skenario menggunakan dataset, arsitektur model, serta parameter pelatihan yang sama. Selanjutnya, kinerja model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC*, serta dianalisis menggunakan *confusion matrix* untuk memahami distribusi kesalahan klasifikasi pada masing-masing kelas.

3.1 Pembagian Data

Dataset akhir yang digunakan dalam penelitian ini berjumlah 3.863 data teks, yang telah melalui tahap pra-pemrosesan dan penghapusan data duplikat sebagaimana dijelaskan pada metodologi penelitian. Dataset tersebut dibagi menjadi tiga subset data menggunakan metode *stratified split*, yaitu 80% digunakan sebagai data latih, 10% untuk data validasi, dan 10% sisanya sebagai data uji. Secara rinci, pembagian data menghasilkan 3.090 data latih, 386 data validasi, dan 387 data uji. Distribusi ini ditunjukkan pada Tabel 4.

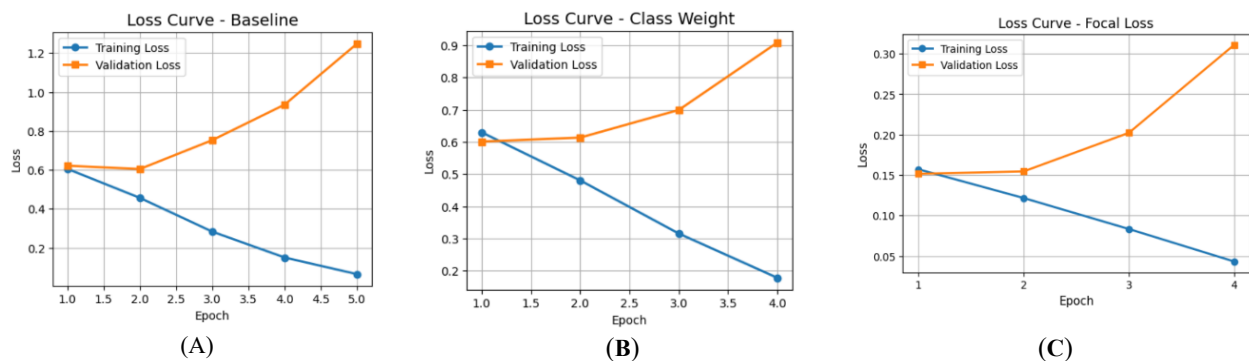
Tabel 4. Hasil Pembagian Data

Subset Data	Jumlah
Data Latih	3.090
Data Validasi	386
Data Uji	387

Penggunaan *stratified split* bertujuan untuk menjaga proporsi distribusi kelas pada setiap subset data agar tetap mencerminkan kondisi dataset asli yang tidak seimbang. Pendekatan ini memiliki peran utama dalam meminimalkan bias evaluasi sekaligus menjamin analisis kinerja model, khususnya pada kelas minoritas, dilakukan secara representatif.

3.2 Hasil Pelatihan Model

Berdasarkan Gambar 3, ketiga skenario pelatihan menunjukkan bahwa *training loss* berkurang seiring bertambahnya jumlah *epoch*, yang menunjukkan bahwa model *IndoBERT* sudah memenuhi pola dalam data latih dengan baik. Namun, dalam semua skenario yang dianalisis, terjadi peningkatan nilai *validasi loss* setelah beberapa *epoch* awal, yang menunjukkan adanya tanda-tanda *overfitting*.



Gambar 3. TRAINING LOSS (a) Base Line, (b) class weight, dan (c) focal loss

Pada *baseline*, terjadi kenaikan yang cukup besar pada *validasi loss* meskipun *training loss* terus berkurang. Skenario *class weight* dan *focal loss* menunjukkan pola yang hampir sama, yaitu peningkatan *validasi loss* yang masih bisa dikendalikan, meskipun belum mampu menghilangkan indikasi *overfitting*. Meskipun metode mengatasi ketidakseimbangan kelas sudah digunakan, kemampuan model dalam menggeneralisasi ke data validasi masih kurang baik.

Tabel 5. Hasil Pelatihan IndoBERT Pada Data Validasi

Metode Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
IndoBERT Baseline	0.7752	0.7752	0.7752	0.7737	0.8528
IndoBERT + Class Weight	0.7468	0.7644	0.7468	0.7467	0.8342
IndoBERT + Focal Loss	0.7287	0.7414	0.7287	0.7291	0.8188

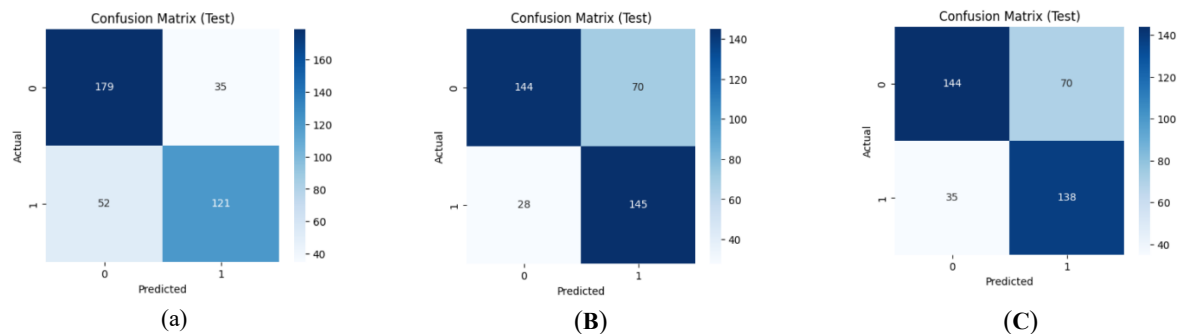
Berdasarkan tabel 5, model *IndoBERT baseline* menunjukkan hasil yang terbaik dengan akurasi 0.7752, *F1-score* 0.7737, dan *ROC-AUC* 0.8528, lebih baik dibandingkan metode *class weight* dan *focal loss*. Penerapan *class weight* dan *focal loss* justru mengakibatkan penurunan di semua metrik evaluasi, yang menunjukkan bahwa teknik penyesuaian distribusi kelas tidak memberikan peningkatan yang berarti pada dataset yang digunakan dalam penelitian ini. Hasil ini menunjukkan bahwa menggunakan fungsi *Cross-Entropy* biasa dengan *optimizer AdamW* sudah cukup bagus dalam menghadapi karakteristik data, sehingga skenario *baseline* dipilih sebagai model terbaik berdasarkan pengevaluasian secara keseluruhan.

3.3 Hasil Pengujian Model

Evaluasi kinerja model dilakukan dengan menggunakan data uji untuk mengevaluasi kemampuan model dalam mengklasifikasikan teks indikasi depresi yang ditulis dalam bahasa Indonesia. Pengujian dilakukan pada tiga skenario model, yaitu *IndoBERT baseline*, *IndoBERT* dengan penerapan *class weight*, dan *IndoBERT* dengan penerapan *focal loss*. Evaluasi dilakukan dengan menggunakan beberapa matrik seperti *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC*. Selain itu, *confusion matrix* digunakan untuk memahami bagaimana kesalahan klasifikasi terdistribusi di setiap kelas yang semuanya dianalisis dari data validasi dan data uji.

3.3.1 Hasil Confusion Metrik Pada Data Uji

untuk menganalisis distribusi tersebut kesalahan klasifikasi pada masing-masing model, digunakan confusion yang menggambarkan perbandingan antara label aktual dan prediksi. Hasil confusion matrik untuk setiap skenario pelatihan ditunjukkan pada Gambar 4.



Gambar 4. Confusion Matrix (a) Base Line, (b) class weight, dan (c) focal loss

Berdasarkan Gambar 4, memiliki karakteristik setiap skenario yang berbeda dalam mengklasifikasikan data. Pada Gambar 4(A), yaitu model *baseline*, diperoleh jumlah *True Negative (TN)* sebanyak 179 dan *True Positive (TP)* sebanyak 121, yang menunjukkan bahwa model tersebut sudah cukup bagus dalam mengklasifikasikan kedua kelas, yaitu negatif dan positif, secara akurat. Namun, terdapat 35 kasus *False Positive (FP)* dan 52 *False Negative (FN)*. Jumlah *FN* yang lebih banyak daripada *FP* menunjukkan bahwa model lebih sering tidak mampu mendeteksi kelas positif daripada salah mengklasifikasikan kelas negatif sebagai positif. Ini menunjukkan bahwa meskipun akurasi secara keseluruhan bagus, model masih cenderung melewatkan beberapa data pada kelas positif.

Pada Gambar 4(B), yaitu model *class weight*, diperoleh nilai *TN* sebanyak 144 dan *TP* sebanyak 145. Jika dibandingkan dengan *baseline*, nilai *TP* meningkat, artinya kemampuan mendeteksi kelas positif kini lebih baik. Namun, nilai *FP* naik menjadi 70, sedangkan *FN* turun menjadi 28. Penurunan *FN* menunjukkan bahwa memberi bobot pada kelas tertentu berhasil mengurangi kesalahan dalam mengenali kelas positif, namun akibatnya adalah meningkatkan kesalahan *FP*, sehingga model lebih banyak mengklasifikasikan kelas negatif sebagai kelas positif.

Sementara itu, pada Gambar 4(C), yaitu model yang menggunakan *focal loss*, diperoleh nilai *TN* sebanyak 144 dan *TP* sebanyak 138. Nilai *FN* sebesar 35 menunjukkan ada peningkatan dibandingkan nilai awal, meskipun masih lebih tinggi dibandingkan *class weight*. Sementara itu, *FP* tetap berada di *focal loss* cukup baik dalam meningkatkan kemampuan mengenali kelas positif dibandingkan dengan metode *baseline*, tetapi masih menghasilkan jumlah prediksi positif palsu yang cukup tinggi. Secara umum, *focal loss* mampu menyeimbangkan antara mengurangi kesalahan klasifikasi jenis salah (*FN*) dan meningkatkan kesalahan klasifikasi jenis benar (*FP*), namun belum bisa mengungguli performa *baseline* secara keseluruhan.

3.3.2 Perbandingan Kinerja Antar Model Pada Data Uji

Berdasarkan hasil pengujian model yang dirangkum pada Tabel 6, setiap skenario pelatihan *IndoBERT* menunjukkan karakteristik performa yang berbeda dalam mengklasifikasikan teks indikasi depresi berbahasa Indonesia. Evaluasi dilakukan dengan menggunakan *matrix accuracy*, *precision*, *recall*, *F1-score*, dengan *weighted* serta *ROC-AUC* agar bisa menilai performa model secara lengkap, terutama pada kondisi dataset yang tidak seimbang.

Tabel 6. Hasil Pelatihan *IndoBERT* Pada Data Uji

Metode Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
<i>IndoBERT</i> Baseline	0.7752	0.7752	0.7752	0.7737	0.8528
<i>IndoBERT</i> + Class Weight	0.7468	0.7644	0.7468	0.7467	0.8342
<i>IndoBERT</i> + Focal Loss	0.7287	0.7414	0.7287	0.7291	0.8188

Dalam skenario *IndoBERT* dasar tanpa penyeimbangan kelas, model menghasilkan akurasi sebesar 0,7726, *weighted F1-score* sebesar 0,7725, dan *ROC-AUC* sebesar 0,8363. hasil ini menunjukkan bahwa model tetap mampu menjaga kinerja klasifikasi yang cukup baik meskipun tidak menerapkan metode khusus untuk mengatasi masalah kelas tidak seimbang. *Confusion matrix* menunjukkan bahwa model cukup konsisten dalam mengenali kedua kelas, dengan jumlah prediksi yang benar untuk kelas Tidak Terindikasi Depresi dan Terindikasi Depresi relatif seimbang. Hal ini mengindikasikan bahwa pendekatan baseline masih mampu memberikan kemampuan generalisasi yang baik pada data uji.

Dalam skenario *IndoBERT* dengan penerapan *class weight*, model menghasilkan akurasi sebesar 0,7519 dan *weighted F1-score* sebesar 0,7526, yang sedikit lebih rendah dibandingkan *baseline*. Namun demikian, nilai *ROC-AUC* naik menjadi 0,8464, yang artinya kemampuan model dalam membedakan distribusi probabilitas antar kelas semakin baik. *Confusion matrix* memperlihatkan peningkatan *recall* pada kelas minoritas Terindikasi Depresi, yang menunjukkan bahwa pemberian bobot kelas berhasil meningkatkan sensitivitas model terhadap data minoritas. Meskipun demikian, peningkatan kesalahan klasifikasi pada kelas mayoritas menyebabkan penurunan akurasi keseluruhan, sehingga menunjukkan adanya *trade-off* antara sensitivitas terhadap kelas minoritas dan kestabilan performa global model.

Sementara itu, skenario *IndoBERT* dengan penerapan *focal loss* menghasilkan *accuracy* sebesar 0,7597 dengan *weighted F1-score* sebesar 0,7596 dan *ROC-AUC* sebesar 0,8241. Nilai tersebut menunjukkan performa yang berada di antara *baseline* dan *class weight*. *Confusion matrix* menunjukkan bahwa *focal loss* mampu meningkatkan keseimbangan prediksi antara kedua kelas tanpa menyebabkan penurunan performa yang terlalu signifikan. Ini menunjukkan bahwa *focal loss* mampu menekankan proses belajar pada sampel yang sulit dikelompokkan, sekaligus mempertahankan keseimbangan performa antar kelas.

Temuan penelitian ini menunjukkan bahwa penerapan metode penanganan ketidakseimbangan kelas tidak selalu menghasilkan peningkatan performa secara langsung pada seluruh matrik evaluasi. Model baseline justru menghasilkan nilai *accuracy* dan *weighted F1-score* tertinggi, sementara *class weight* menghasilkan nilai *ROC-AUC* terbaik serta peningkatan sensitivitas terhadap kelas minoritas. *Focal loss* menunjukkan performa kompromi yang relatif seimbang antara kedua pendekatan tersebut.

Secara keseluruhan, hasil perbandingan antar skenario menegaskan bahwa pemilihan metode penanganan ketidakseimbangan kelas perlu mempertimbangkan tujuan evaluasi model secara spesifik. Jika fokus utama adalah akurasi keseluruhan, pendekatan *baseline* dapat menjadi pilihan yang optimal. Namun, apabila tujuan penelitian adalah meningkatkan kemampuan deteksi kelas minoritas atau meningkatkan kualitas pemisahan probabilitas antar kelas, maka penerapan *class weight* atau *focal loss* dapat memberikan keuntungan tertentu. Dengan demikian, evaluasi model klasifikasi teks indikasi depresi tidak dapat hanya didasarkan pada satu metrik tunggal, melainkan harus mempertimbangkan keseimbangan antara akurasi, sensitivitas kelas minoritas, dan kemampuan diskriminatif model secara menyeluruh.

3.4 Pembahasan

Hasil penelitian menunjukkan bahwa model *IndoBERT* baseline memberikan performa terbaik pada *matrik accuracy* dan *weighted F1-score*, yang mengindikasikan bahwa model mampu melakukan generalisasi dengan baik meskipun tanpa penanganan khusus terhadap ketidakseimbangan kelas. Hal ini menunjukkan bahwa representasi kontekstual yang dihasilkan oleh model *transformer* yang telah dilatih sebelumnya cukup kuat dalam memahami karakteristik teks berbahasa Indonesia, termasuk dalam mendeteksi indikasi depresi. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa model berbasis BERT memiliki kemampuan unggul dalam tugas klasifikasi teks karena mampu menangkap konteks semantik secara lebih mendalam dibandingkan metode konvensional. Sementara itu, penerapan metode *class weight* menunjukkan peningkatan pada nilai *ROC-AUC* serta *recall* pada kelas minoritas, yang menandakan bahwa model menjadi lebih sensitif dalam mendeteksi teks yang mengandung indikasi depresi. Namun, akibat meningkatnya kesalahan klasifikasi ini diikuti dengan penurunan *accuracy* secara keseluruhan akibat meningkatnya kesalahan klasifikasi pada kelas mayoritas. Fenomena ini menunjukkan adanya *trade-off* antara kemampuan mendeteksi kelas minoritas dan kestabilan performa global model, yang juga banyak ditemukan dalam penelitian terkait penanganan ketidakseimbangan kelas pada tugas klasifikasi teks. Adapun penerapan *focal loss* menghasilkan performa yang relatif seimbang di antara kedua pendekatan tersebut, dengan kemampuan yang cukup baik dalam menekankan pembelajaran

pada sampel yang sulit diklasifikasikan tanpa menyebabkan penurunan performa yang terlalu signifikan. Hal ini menunjukkan bahwa *focal loss* dapat menjadi alternatif dalam mengenai ketidakseimbangan kelas dengan pendekatan yang lebih adaptif terhadap distribusi kesalahan model. Secara keseluruhan, hasil ini menunjukkan bahwa tidak terdapat satu metode yang secara konsisten unggul pada seluruh metrik evaluasi. Oleh karena itu, pemilihan metode pelatihan perlu disesuaikan dengan tujuan penggunaan model, apakah lebih berfokus pada keseluruhan, peningkatan sensitivitas terhadap kelas minoritas, atau kemampuan diskriminatif model dalam membedakan probabilitas antar kelas. Temuan ini juga memperkuat pentingnya penggunaan berbagai metrik evaluasi secara bersamaan dalam menilai kinerja model klasifikasi teks, terutama pada kasus dengan distribusi data yang tidak seimbang seperti deteksi indikasi depresi.

4. KESIMPULAN

Berdasarkan hasil penelitian mengenai klasifikasi teks indikasi depresi berbahasa Indonesia menggunakan model *IndoBERT* dengan tiga skenario pelatihan, yaitu *baseline*, *class weight*, dan *focal loss*, dapat disimpulkan bahwa setiap pendekatan memiliki karakteristik performa yang berbeda ketika dievaluasi menggunakan data *testing*. Model *IndoBERT baseline* menghasilkan performa terbaik dalam hal *accuracy* mencapai 0,7726 dan *weighted F1-score* sebesar 0,7725, menunjukkan kemampuan model dalam generalisasi yang stabil tanpa penerapan metode penanganan ketidakseimbangan kelas secara khusus. Model ini juga menunjukkan distribusi prediksi yang relatif seimbang antara kelas Tidak Terindikasi Depresi dan Terindikasi Depresi berdasarkan hasil *confusion matrix*. Sementara itu, penerapan *class weight* berhasil memberikan nilai *ROC-AUC* 0,8464, yang menunjukkan kemampuan model lebih baik dalam membedakan probabilitas antar kelas. Pendekatan ini juga meningkatkan sensitivitas terhadap kelas minoritas Terindikasi Depresi, meskipun terjadi penurunan *accuracy* keseluruhan menjadi 0,7519 akibat meningkatnya kesalahan klasifikasi pada kelas mayoritas. Adapun penerapan *focal loss* menghasilkan performa yang berada di antara kedua pendekatan lainnya dengan *accuracy* sebesar 0,7597, *weighted F1-score* sebesar 0,7596, dan *ROC-AUC* sebesar 0,8241. Hasil ini menunjukkan bahwa *focal loss* mampu menjaga keseimbangan performa antara stabilitas model dan sensitivitas terhadap sampel yang sulit diklasifikasikan, sehingga memberikan kompromi yang relatif seimbang antara kedua metode lainnya. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa metode penanganan ketidakseimbangan kelas tidak selalu menghasilkan peningkatan performa global model dalam seluruh metrik evaluasi. Model *baseline* justru memberikan performa terbaik pada *matrix accuracy* dan *weighted F1-score*, sedangkan *class weight* unggul dalam kemampuan diskriminatif model yang tercermin pada nilai *ROC-AUC*. Dengan demikian, pemilihan metode pelatihan perlu disesuaikan dengan tujuan penggunaan model, baik untuk akurasi keseluruhan, sensitivitas terhadap kelas minoritas, maupun kemampuan pemisahan probabilitas. Selain itu, evaluasi model tidak cukup hanya mengandalkan metrik pelatihan seperti *training loss*, tetapi harus mempertimbangkan performa pada data *testing* sebagai indikator utama kemampuan generalisasi. Penelitian ini masih memiliki keterbatasan, antara lain penggunaan dataset dari satu sumber dengan kontribusi kelas yang tidak seimbang, serta eksperimen yang terbatas pada satu arsitektur transformer lainnya. Oleh karena itu, penelitian selanjutnya disarankan untuk melakukan perbandingan dengan berbagai arsitektur model serta menerapkan pendekatan *Explainable Artificial Intelligence (XAI)* guna meningkatkan transparansi dan kepercayaan dalam pengambilan keputusan model, terutama pada kasus klasifikasi teks yang sensitif seperti deteksi depresi.

REFERENCES

- [1] M. Y. Dhinora and E. Mailoa, "Analisa Tweet Mahasiswa untuk Deteksi Gejala Depresi dengan Penerapan Natural Language Processing," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 6, no. 2, pp. 1193–1211, 2025, doi: 10.63447/jimik.v6i2.1405.
- [2] G. F. Situmorang and R. Purba, "Deteksi Potensi Depresi dari Unggahan Media Sosial X Menggunakan IndoBERT," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, pp. 649–661, Sep. 2024, doi: 10.47065/bits.v6i2.5496.
- [3] A. G. S. Moeslim, E. Firmansyah, and B. Sutara, "Perbandingan Kinerja XGBoost dan IndoBERT untuk Klasifikasi Teks Kesehatan Bahasa Indonesia," *Data Sciences Indonesia (DSI)*, vol. 5, no. 2, pp. 62–72, Dec. 2025, doi: 10.47709/dsi.v5i2.7281.
- [4] M. B. M. Amin *et al.*, "Deteksi Spam Berbahasa Indonesia Berbasis Teks Menggunakan Model Bert," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1291–1302, Dec. 2024, doi: 10.25126/jtiik.2024118121.
- [5] M. Fitri, J. S. Nurkhotimah, F. N. Ihsan, K. Amaliah, and D. Rofianto, "Model Ensemble untuk Deteksi Depresi di Twitter Berbahasa Indonesia," *Jurnal Sifo Mikroskil*, vol. 26, no. 2, Oct. 2025, doi: 10.55601/jsm.v26i2.1812.
- [6] C. Sintiya, G. H. Hutagaol, D. Bate'e, and S. Irviantina, "Evaluasi Teknik Resampling untuk Class Balancing dalam Analisis Sentimen Kesehatan Mental Berbasis Bi-LSTM," *Jurnal Sifo Mikroskil*, vol. 26, no. 2, pp. 257–274, Oct. 2025, doi: 10.55601/jsm.v26i2.1799.
- [7] H. F. Karim and A. P. Wibowo, "Kinerja Metode Fine-Tuning IndoBERT untuk Klasifikasi Emosi Multi-Kelas pada Teks Informal Bahasa Indonesia," *Bulletin of Computer Science Research*, vol. 6, no. 1, pp. 63–74, 2025, doi: 10.47065/bulletincsr.v6i1.850.
- [8] E. P. Adamansyah and A. Yudhistira, "Evaluasi Opini Publik di Media Sosial X terhadap Kebijakan Pajak Pertambahan Nilai 12% di Indonesia Menggunakan Naive Bayes dan Decision Tree," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 3, pp. 831–843, Mar. 2025, doi: 10.52436/1.jpti.710.
- [9] A. A. Pangestu and A. R. Purnajaya, "Perbandingan Performa Arsitektur Machine Learning untuk Deteksi Dini Depresi Berbasis Natural Language Processing dalam Bahasa Indonesia," *J-SIGN (Journal of Informatics, Information System, and Artificial Intelligence)*, vol. 3, no. 2, pp. 93–104, Oct. 2025, doi: 10.24815/j-sign.v3i2.49873.

- [10] Z. Darojah, R. Susetyoko, N. Ramadijanti, P. Elektronika Negeri Surabaya, and P. Korespondensi, “Strategi Penanganan Imbalance Class Pada Model Klasifikasi Penerimaan Kartu Indonesia Pintar Kuliah Berbasis Neural Network Menggunakan Kombinasi SMOTE dan ENN,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 10, no. 2, pp. 457–466, 2023, doi: 10.25126/jtiik.2023106480.
- [11] I. A. Rahma and L. H. Suadaa, “Penerapan Text Augmentation untuk Mengatasi Data yang Tidak Seimbang pada Klasifikasi Teks Berbahasa Indonesia,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 6, pp. 1329–1340, Dec. 2023, doi: 10.25126/jtiik.2023107325.
- [12] M. Fajar Ramadhan, Febriyanti Panjaitan, Winarnie, Hery Oktafiandi, and Yohanes, “Analisis Sentimen Kutipan Media Sosial Berbahasa Indonesia Menggunakan Convolutional Neural Network,” *Jurnal Komputer, Informasi dan Teknologi*, vol. 5, no. 2, pp. 17–26, Dec. 2025, doi: 10.53697/jkomitek.v6i1.3627.
- [13] E. Sutoyo and M. C. Permana, “Enhancing telemedicine service quality through sentiment analysis of user review dataset in Indonesia,” *Data Brief*, vol. 61, p. 111878, Aug. 2025, doi: 10.1016/j.dib.2025.111878.
- [14] T. Jia, R. de Vries, Z. Kapelan, T. H. M. van Emmerik, and R. Taormina, “Detecting floating litter in freshwater bodies with semi-supervised deep learning,” *Water Res.*, vol. 266, p. 122405, Nov. 2024, doi: 10.1016/j.watres.2024.122405.
- [15] M. J. Faisti, R. Hadapiningradja Kusumodestoni, G. Wahyu, and N. Wibowo, “Mental Health Classification Using Naïve Bayes and Random Forest Algorithms,” *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 4, pp. 1740–1750, 2025, [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [16] R. Oktafiani, A. Hermawan, and D. Avianto, “Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning,” *Jurnal Sains dan Informatika*, pp. 19–28, Jun. 2023, doi: 10.34128/jsi.v9i1.622.
- [17] M. R. Manoppo *et al.*, “Analisis Sentimen Publik Di Media Sosial Terhadap Kenaikan PPN 12% di Indonesia Menggunakan IndoBERT,” *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 2, pp. 152–163, May 2025, doi: 10.69916/jkbt.v4i2.322.
- [18] I. Mursidah, R. Sanjaya, B. Yulianto, D. Sweetania, and P. Sularsih, “Klasifikasi Sentimen Google Play Store Aplikasi ChatGPT Berbahasa Indonesia Berbasis IndoBERT,” *Jurnal Minfo Polgan*, vol. 14, no. 2, pp. 3349–3359, Dec. 2025, doi: 10.33395/jmp.v14i2.15751.
- [19] A. F. Hidayatullah, R. A. Apong, D. T. C. Lai, and A. Qazi, “Corpus creation and language identification for code-mixed Indonesian-Javanese-English Tweets,” *PeerJ Comput. Sci.*, vol. 9, p. e1312, Jun. 2023, doi: 10.7717/peerj-cs.1312.
- [20] N. N. A. Aryanti and O. Suria, “Analisis Sentimen Terhadap Pemutusan Hubungan Kerja di Indonesia : Komparasi IndoBERT dengan SVM, Random Forest, dan Decision Tree Dengan Optimasi TF - IDF,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 1158–1176, Jul. 2025, doi: 10.36341/rabit.v10i2.6364.
- [21] F. Indriani, R. A. Nugroho, M. R. Faisal, and D. Kartini, “Comparative Evaluation of IndoBERT, IndoBERTweet, and mBERT for Multilabel Student Feedback Classification,” *Jurnal RESTI*, vol. 8, no. 6, pp. 748–757, Dec. 2024, doi: 10.29207/resti.v8i6.6100.
- [22] D. Ramadhan Alfinsyah and B. Pulu Hartato, “Evaluating the Impact of Random Over Sampling on IndoBERT Performance for Indonesian Sentiment Analysis,” *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 6, pp. 3270–3282, 2025, [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [23] Yohanes Anton Nugroho and Hotma Antoni Hutahaean, “Pengembangan Model Deep Learning LSTM dan CNN untuk Peramalan Penjualan Sepeda Motor di Indonesia,” *Jupiter: Publikasi Ilmu Keteknikan Industri, Teknik Elektro dan Informatika*, vol. 3, no. 2, pp. 94–104, Mar. 2025, doi: 10.61132/jupiter.v3i2.795.
- [24] M. N. Razali, N. Arbaiy, P.-C. Lin, and S. Ismail, “Optimizing Multiclass Classification Using Convolutional Neural Networks with Class Weights and Early Stopping for Imbalanced Datasets,” *Electronics (Basel)*, vol. 14, no. 4, p. 705, Feb. 2025, doi: 10.3390/electronics14040705.
- [25] A. I. Faiq, M. R. S. Sanjaya, D. R. Indah, and E. L. Ruskan, “Hybrid Fine-Tuning IndoBERT dan Ensemble TF-IDF Logistic Regression Untuk Analisis Sentimen Ulasan Aplikasi Zalora,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 11, no. 1, pp. 324–336, Jan. 2026, doi: 10.36341/rabit.v11i1.6924.
- [26] A. Kunaefi, Z. Abidin, and R. Kusumawati, “Klasifikasi Berita Hoaks Bahasa Indonesia Menggunakan IndoBERT Fine - Tuning Dengan Pendekatan Focal Loss Pada Data Tidak Seimbang,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1706–1714, May 2025, doi: 10.29100/jupi.v10i2.7811.
- [27] F. Alzami *et al.*, “Bayesian-Optimized CLAHE for Enhanced Drowsiness Detection in Low-Light Conditions Using Time-Distributed MobileNetV2-GRU Architecture,” *Statistics, Optimization and Information Computing*, vol. 15, no. 1, pp. 274–294, Dec. 2025, doi: 10.19139/soic-2310-5070-3024.
- [28] V. W. Handayani, M. S. Margareth Amiatun Ruth, R. Rulaningtyas, M. R. Caesarardhi, B. A. Yudhantorro, and A. Yudianto, “Development and evaluation of a convolutional neural network model for sex prediction using cephalometric radiographs and cranial photographs,” *BMC Med. Imaging*, vol. 25, no. 1, p. 348, Aug. 2025, doi: 10.1186/s12880-025-01892-x.