Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

Studi Komparasi IndoBERT dan IndoBERTweet dengan Full Fine Tuning dan LoRA untuk Klasifikasi Emosi pada Tweet Bertopik GERD

Luthfia Jayyida Ainaya Fatiha, Junta Zeniarja*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1]111202214609@mhs.dinus.ac.id, [2,*]junta@dsn.dinus.ac.id
Email Penulis Korespondensi: junta@dsn.dinus.ac.id
Submitted: **26/05/2026**; Accepted: **30/06/2026**; Published**: 30/06/2026**

Abstrak−Penelitian ini membandingkan performa IndoBERT dan IndoBERTweet dalam klasifikasi emosi pada tweet berbahasa Indonesia yang bertopik Gastroesophageal Reflux Disease (GERD). Istilah tweet bertopik GERD digunakan karena data berasal dari unggahan media sosial X, sehingga status klinis pengguna tidak dapat dipastikan. Dataset awal berjumlah 1.389 data dan setelah proses validasi, pembersihan teks, penghapusan data kosong, penyaringan label, serta deduplikasi, diperoleh 1.343 tweet yang terdiri atas lima kelas emosi, yaitu angry, fear, joy, sad, dan netral. Eksperimen dilakukan menggunakan Python pada Google Colab dengan GPU Tesla T4, serta memanfaatkan Pandas, PyTorch, Hugging Face Transformers, PEFT, scikit-learn, Matplotlib, dan Seaborn. Model dievaluasi menggunakan Stratified 5-Fold Cross Validation, data augmentation berbasis substitusi leksikal pada data training setiap fold, early stopping, dan Focal Loss untuk menangani ketidakseimbangan kelas. Hasil menunjukkan bahwa full fine-tuning menghasilkan performa terbaik, terutama IndoBERTweet dengan accuracy 0,5703 dan macro F1-score 0,5629. IndoBERT full fine-tuning memperoleh accuracy 0,5145 dan macro F1-score 0,5034. Sebaliknya, IndoBERT + LoRA hanya memperoleh macro F1-score 0,2412, sedangkan IndoBERTweet + LoRA memperoleh macro F1-score 0,2227. Dengan demikian, LoRA belum efektif untuk dataset kecil dan tidak seimbang pada domain emosi kesehatan ini, meskipun hanya melatih kurang dari 1% parameter model. Temuan ini menegaskan bahwa efisiensi parameter tidak selalu sejalan dengan peningkatan performa klasifikasi pada teks media sosial yang informal, ambigu, dan memiliki tumpang tindih emosi.

Kata Kunci: GERD; IndoBERT; IndoBERTweet; Klasifikasi Emosi; LoRA; Full Fine-Tuning; Media Sosial X

Abstract−This study compares IndoBERT and IndoBERTweet for emotion classification on Indonesian tweets related to Gastroesophageal Reflux Disease (GERD). The term GERD-related tweets is used because the data were collected from social media X, and the clinical status of each user cannot be verified. The initial dataset contained 1,389 records. After validation, text cleaning, empty-text removal, label filtering, and deduplication, 1,343 tweets remained and were manually assigned into five emotion classes: angry, fear, joy, sad, and neutral. The experiments were conducted in Python on Google Colab with a Tesla T4 GPU using Pandas, PyTorch, Hugging Face Transformers, PEFT, scikit-learn, Matplotlib, and Seaborn. The models were evaluated using Stratified 5-Fold Cross Validation, lexical substitution-based data augmentation on the training data of each fold, early stopping, and Focal Loss to address class imbalance. The results show that full fine-tuning achieved the best performance. IndoBERTweet obtained the highest accuracy of 0.5703 and macro F1-score of 0.5629, while IndoBERT achieved an accuracy of 0.5145 and a macro F1-score of 0.5034. In contrast, IndoBERT + LoRA only achieved a macro F1-score of 0.2412, and IndoBERTweet + LoRA achieved 0.2227. These findings indicate that LoRA was not effective for this small and imbalanced health-emotion dataset, although it trained less than 1% of the model parameters. Parameter efficiency therefore did not translate into better classification performance in informal and semantically overlapping social media text.

Keywords: GERD; IndoBERT; IndoBERTweet; Emotion Classification; LoRA; Full Fine-Tuning; Social Media X

1. PENDAHULUAN

Gastro-esophageal Reflux Disease (GERD) merupakan gangguan kronis pada saluran pencernaan yang terjadi akibat refluks asam lambung berulang ke esofagus dan dapat menimbulkan gejala seperti *heartburn*, regurgitasi, disfagia, nyeri dada non-kardiak, mual, serta gangguan tidur [1]. Secara global, prevalensi GERD dilaporkan berkisar antara 7,4% hingga 19,6% dan dipengaruhi oleh pola makan, gaya hidup, obesitas, serta faktor psikologis seperti stres [2]. *Global Burden of Disease Study* 2021 juga menunjukkan peningkatan jumlah kasus GERD dari sekitar 450,76 juta pada tahun 1990 menjadi lebih dari 825 juta pada tahun 2021, dan diproyeksikan terus meningkat hingga tahun 2050 [3].

GERD tidak hanya berdampak secara fisik, tetapi juga berkaitan dengan kondisi psikologis penderitanya. Henning et al. (2024) menunjukkan bahwa pasien dengan gejala refluks memiliki tingkat kecemasan yang tinggi, dimana 44,1% responden menunjukkan skor kecemasan di atas batas normal berdasarkan *Hospital Anxiety and Depression Scale*. Stres juga dapat mempengaruhi sistem pencernaan melalui peningkatan sekresi asam lambung dan perubahan motilitas gastrointestinal, sehingga berpotensi memperburuk gejala refluks [4]. Kondisi ini membuat ekspresi emosi penderita GERD menjadi penting untuk dianalisis, terutama pada media sosial yang sering digunakan untuk menyampaikan pengalaman kesehatan dan keluhan emosional secara terbuka [5].

Media sosial seperti *Twitter* atau X menyediakan data teks pendek yang relevan untuk memahami ekspresi emosi karena pengguna sering menyampaikan suasana hati, keluhan fisik, dan pengalaman personal secara spontan [6]. Penelitian Mayor et al. (2022) juga menunjukkan bahwa unggahan *Twitter* dapat merepresentasikan kondisi emosional individu maupun kelompok [7]. Namun, teks media sosial memiliki tantangan khusus karena menggunakan bahasa informal, singkatan, bahasa gaul, emoji, serta konteks implisit. Oleh karena itu, *Natural Language Processing* (NLP) diperlukan untuk membantu komputer memahami dan mengklasifikasikan makna emosi dalam teks tersebut [8].

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

Perkembangan *pre-trained language models* berbasis *Transformer* telah meningkatkan performa berbagai tugas NLP, termasuk analisis sentimen dan klasifikasi emosi [9]. Salah satu model yang banyak digunakan adalah *Bidirectional Encoder Representations from Transformers* (BERT), yang dirancang untuk memahami konteks kata secara *bidirectional* sehingga mampu menangkap hubungan makna dalam kalimat [10]. Dalam bahasa Indonesia, IndoBERT dikembangkan menggunakan korpus bahasa Indonesia umum, sedangkan IndoBERTweet dirancang lebih spesifik untuk teks media sosial seperti *Twitter* [11], [12]. Perbedaan karakteristik data pelatihan ini membuat kedua model relevan untuk dibandingkan dalam klasifikasi emosi pada teks media sosial berbahasa Indonesia.

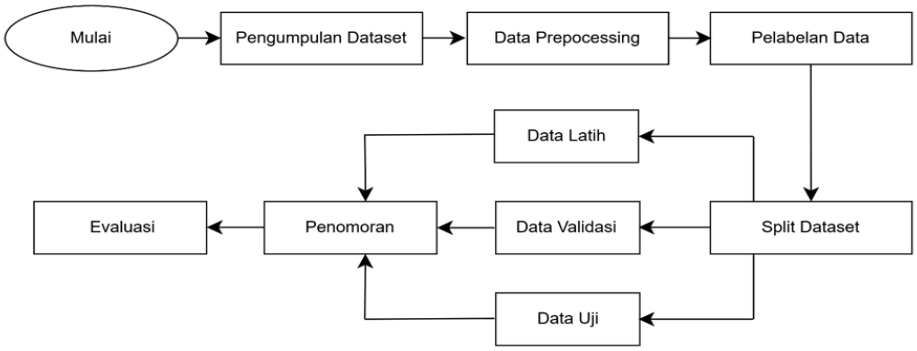
Sejumlah penelitian menunjukkan bahwa model berbasis BERT memiliki performa yang baik dalam analisis sentimen dan klasifikasi emosi. IndoBERT pernah digunakan untuk analisis sentimen Pemilu 2024 di *Twitter* dengan akurasi 60%, sedangkan IndoBERTweet memperoleh akurasi 80% pada ulasan TikTok berbahasa Indonesia [13], [14]. Penelitian lain juga menunjukkan performa IndoBERTweet yang tinggi pada deteksi emosi dan risiko depresi di media sosial [15], [16]. Meskipun demikian, penelitian tersebut masih banyak berfokus pada topik umum, politik, ulasan pengguna, dan kesehatan mental. Kajian klasifikasi emosi pada penyakit fisik seperti GERD masih terbatas, terutama yang membandingkan IndoBERT dan IndoBERTweet pada domain kesehatan berbasis media sosial.

Berdasarkan kesenjangan tersebut, penelitian ini membandingkan IndoBERT dan IndoBERTweet dalam klasifikasi emosi penderita GERD pada media sosial menggunakan *full fine-tuning* dan *Parameter-Efficient Fine-Tuning* (PEFT) berbasis *Low-Rank Adaptation* (LoRA). Pemilihan LoRA dalam penelitian ini tidak diposisikan sebagai metode yang pasti lebih unggul, melainkan sebagai pendekatan pembanding untuk menguji apakah adaptasi parameter terbatas tetap efektif pada *dataset* kecil dan tidak seimbang. Dengan jumlah data 1.343 *tweet*, penelitian ini secara metodologis menilai batas efektivitas LoRA dalam konteks klasifikasi emosi kesehatan berbasis media sosial. Oleh karena itu, kontribusi penelitian ini terletak pada perbandingan performa antara *full fine-tuning* dan LoRA, sekaligus analisis kesesuaian strategi *fine-tuning* terhadap karakteristik *dataset* kecil, tidak seimbang, dan berbahasa informal.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini diawali dengan pengumpulan data melalui proses *crawling* pada media sosial X untuk memperoleh unggahan berbahasa Indonesia yang berkaitan dengan GERD. Data kemudian diproses melalui tahap *preprocessing*, pelabelan manual, dan validasi label. Eksperimen dilakukan menggunakan *Python* pada lingkungan *Google Colab* dengan dukungan *GPU*. Beberapa *framework* dan *library* yang digunakan meliputi *Pandas* dan *NumPy* untuk pengolahan data, *Scikit-learn* untuk *Stratified 5-Fold Cross Validation* dan evaluasi metrik, *PyTorch* sebagai *backend* pelatihan model, serta *Hugging Face Transformers* dan PEFT untuk implementasi IndoBERT, IndoBERTweet, *full fine-tuning*, dan LoRA. Evaluasi model dilakukan menggunakan *early stopping*, *Focal Loss*, *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, dan *error analysis*.



Gambar 1. Alur penelitian

2.2 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan menggunakan teknik web crawling terhadap unggahan pengguna media sosial X yang berkaitan dengan Gastroesophageal Reflux Disease (GERD). Proses crawling dilakukan pada periode Oktober-Januari menggunakan kata kunci “GERD” untuk memperoleh data berupa teks berbahasa Indonesia yang memuat pengalaman, ekspresi, dan keluhan pengguna terkait penyakit tersebut. Seluruh data hasil crawling disimpan dalam format Comma Separated Values (CSV). Berdasarkan hasil pemeriksaan data, jumlah data mentah yang diperoleh sebanyak 1.1440 tweet. Setelah dilakukan validasi, penghapusan data kosong, penyaringan label yang tidak sesuai, dan penghapusan duplikasi, sehingga jumlah data akhir yang digunakan dalam penelitian adalah 1.343 tweet. Penelitian ini hanya menggunakan unggahan yang bersifat publik, tidak mencantumkan identitas pengguna, serta menghapus informasi personal seperti username, mention, dan tautan untuk menjaga privasi dan etika penggunaan data media sosial.

Copyright © 2026 **Author**, Page 656
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

2.3 Preprocessing Data

Preprocessing merupakan tahap awal dalam *text mining* untuk membersihkan dan menyiapkan data teks sebelum proses klasifikasi[17]. Tahap ini dilakukan karena data media sosial umumnya masih mengandung elemen tidak relevan, seperti URL, simbol, angka, dan karakter lain yang dapat mempengaruhi performa model.

Tahap awal dilakukan dengan memuat dataset hasil *crawling* dari media sosial X dalam format CSV menggunakan *library* pandas. Selanjutnya dilakukan *data cleaning* untuk menghapus URL, *mention*, simbol, emoji, dan karakter yang tidak relevan. Setelah itu diterapkan *case folding* untuk mengubah seluruh teks menjadi huruf kecil serta normalisasi kata untuk mengubah bahasa tidak baku dan singkatan menjadi bentuk baku.

Tahap berikutnya adalah tokenisasi untuk memisahkan teks menjadi token kata. Penelitian ini tidak menerapkan *stemming* karena model Transformer menggunakan tokenisasi berbasis *subword* yang mampu memahami struktur morfologis kata secara kontekstual. Tahap terakhir adalah penghapusan data kosong guna menjaga kualitas dataset dan mengurangi potensi bias pada proses pelatihan model.

Tabel 1. Preprocessing Data

Sebelum Preprocessing	Setelah Preprocessing
Sekalinya makan seblak abis itu panas sedada dan perut. Menyala gerd ku	sekalinya makan seblak abis itu panas sedada dan perut menyala gerd ku
@shoubrave Tapi itu indmret kadang udah pada habis duluan dan ngantri nya beneran lama keburu gerd	tapi itu indmret kadang udah pada habis duluan dan ngantri nya beneran lama keburu gerd

2.4 Pelabelan Data

Dataset penelitian terdiri dari lima kelas emosi, yaitu *angry*, *fear*, *joy*, *sad*, dan *netral*. Pelabelan data dilakukan secara manual oleh dua annotator yang memiliki pemahaman terhadap analisis teks bahasa Indonesia dan klasifikasi emosi. Untuk mengurangi subjektivitas, annotator menggunakan pedoman pelabelan yang berisi definisi setiap kelas emosi, yaitu angry, fear, joy, sad, dan netral. Setiap data diberi label berdasarkan makna kalimat, konteks unggahan, serta nuansa emosional yang muncul dalam teks. Hasil pelabelan kemudian melalui proses cross-check antar annotator. Apabila terdapat perbedaan label, keputusan akhir ditentukan melalui diskusi hingga mencapai kesepakatan. Dengan pendekatan ini, dataset yang dihasilkan diharapkan mampu merepresentasikan kondisi emosional pengguna secara lebih akurat, khususnya dalam konteks unggahan terkait penyakit GERD pada media sosial.

2.5 Data Augmentation

Data augmentation digunakan untuk memperkaya variasi data latih dan meningkatkan kemampuan generalisasi model [18]. Pada penelitian ini, augmentasi hanya diterapkan pada data latih di setiap fold untuk mencegah data leakage antara data latih dan data validasi. Teknik augmentasi dilakukan menggunakan substitusi leksikal berbasis sinonim melalui fungsi Python sederhana yang dibuat secara manual. Beberapa kata yang sering muncul dalam konteks keluhan GERD diganti dengan kata lain yang memiliki makna serupa, misalnya “sedih” menjadi “murung” atau “kecewa”, “takut” menjadi “cemas” atau “khawatir”, “marah” menjadi “kesal” atau “emosi”, dan “sakit” menjadi “nyeri” atau “tidak nyaman”. Augmentasi difokuskan pada kelas minoritas agar distribusi data latih menjadi lebih seimbang tanpa mengubah makna utama teks.

2.6 Pembagian Dataset

Penelitian ini menggunakan *Stratified K-Fold Cross Validation* untuk menjaga proporsi distribusi label pada setiap *fold* sehingga evaluasi model menjadi lebih stabil dan representatif. Pada setiap iterasi, satu *fold* digunakan sebagai data uji dan sisanya sebagai data latih. Pendekatan ini dinilai lebih efektif dibandingkan *train-test split* statis karena mampu mengurangi bias evaluasi dan meningkatkan kemampuan generalisasi model pada dataset dengan distribusi kelas yang tidak seimbang.

2.7 Arsitektur Model dan Parameter-Efficient Fine-Tuning

Penelitian ini menggunakan IndoBERT dan IndoBERTweet untuk klasifikasi emosi dengan pendekatan *Parameter-Efficient Fine-Tuning* (PEFT) berbasis *Low-Rank Adaptation* (LoRA). Pendekatan ini hanya melatih parameter tambahan berdimensi rendah tanpa memperbarui seluruh parameter model utama, sehingga lebih efisien dari sisi memori, waktu pelatihan, dan risiko *overfitting* pada dataset terbatas [18]. Selain membandingkan performa model, penelitian ini juga menekankan efisiensi pemrosesan teks medis berbasis media sosial.

2.8 Desain Eksperimen

Penelitian ini menerapkan *stratified 5-fold cross-validation* untuk menghasilkan evaluasi yang lebih stabil dan representatif dengan menjaga proporsi distribusi label pada setiap *fold*. Pada setiap iterasi, satu *fold* digunakan sebagai data validasi dan sisanya sebagai data latih. Selain itu, diterapkan *early stopping* dengan memantau *validation loss* untuk menghentikan pelatihan secara otomatis ketika performa model tidak lagi meningkat, sehingga dapat mengurangi risiko *overfitting*.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

2.9 Fungsi Loss dan Penanganan Ketidakseimbangan Kelas

Ketidakseimbangan distribusi kelas menjadi salah satu tantangan utama dalam klasifikasi emosi pada data GERD karena model cenderung lebih berpihak pada kelas mayoritas. Kondisi ini dapat menyebabkan *accuracy paradox*, yaitu ketika nilai akurasi terlihat tinggi namun kemampuan model dalam mengenali kelas minoritas masih rendah [19]. Untuk mengatasi hal tersebut, penelitian ini menggunakan *Focal Loss* sebagai pengganti *Cross-Entropy Loss*. *Focal Loss* memberikan penekanan lebih besar pada sampel yang sulit diklasifikasikan dan kelas minoritas sehingga diharapkan mampu meningkatkan sensitivitas model serta menghasilkan performa klasifikasi yang lebih seimbang [20]. Formula Focal Loss ditunjukkan pada Persamaan (1).

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \quad (1)$$

Pada Persamaan (1), p_t menunjukkan probabilitas prediksi model terhadap kelas yang benar, α_t menunjukkan bobot kelas untuk menangani ketidakseimbangan data, dan γ merupakan focusing parameter yang mengatur tingkat penekanan terhadap sampel yang sulit diklasifikasikan. Semakin kecil nilai p_t, semakin besar penalti yang diberikan oleh Focal Loss. Dengan demikian, model tidak hanya berfokus pada kelas mayoritas yang mudah dikenali, tetapi juga belajar lebih baik dari sampel kelas minoritas dan sampel yang sulit diprediksi.

2.10 Evaluasi Model

Evaluasi performa model pada penelitian ini dilakukan menggunakan beberapa metrik klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* dalam skala *macro* maupun *weighted* untuk memberikan gambaran yang lebih komprehensif terhadap kemampuan model dalam mengklasifikasikan setiap kelas emosi. Selain itu, digunakan *confusion matrix* untuk mengidentifikasi pola kesalahan prediksi antar kelas sehingga dapat diketahui kelas emosi yang sering tertukar atau sulit dikenali oleh model. Dalam kondisi distribusi data yang tidak seimbang, *macro F1-score* menjadi metrik utama karena mampu merepresentasikan performa model secara lebih adil pada seluruh kelas tanpa dipengaruhi dominasi kelas mayoritas. Dengan demikian, penggunaan *accuracy* saja dinilai belum cukup untuk menggambarkan kemampuan model dalam mengenali seluruh kelas emosi secara proporsional. Hasil evaluasi tidak hanya berfokus pada penyajian nilai metrik, tetapi juga dilengkapi dengan *error analysis* serta interpretasi dari sudut pandang medis untuk memahami karakteristik kesalahan model dan keterkaitannya dengan aspek linguistik maupun klinis dari ekspresi emosi penderita GERD. Dalam penelitian ini, *confusion matrix* digunakan sebagai dasar evaluasi dengan memanfaatkan komponen *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [21]. Hasil yang diperoleh kemudianian dibandingkan dan dianalisis untuk mengevaluasi performa masing-masing model secara lebih mendalam.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Pada persamaan (1), *Accuracy* menggambarkan perbandingan antara prediksi yang benar dengan total prediksi yang dibuat oleh pengklasifikasi yang ada pada data uji.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Persamaan (2) adalah contoh persamaan dari *Precision* atau nilai prediksi positif yang mewakil jumlah prediksi yang benar diantara semua prediksi positif.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Pada persamaan (3) adalah *Recall* atau bisa disebut dengan *sensitivity*, *Recall* mengukur proporsi relatif dari contoh positf yang sudah diklasifikasikan dengan benar dari keseluruhan contoh positif.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

Persamaan (4) dapat dihitung berdasarkan nilai *Precision* dan *Recall.* F1-score memberikan keseimbangan antara *Precision* dan *Recall.*

3. HASIL DAN PEMBAHASAN

Bab ini membahas hasil pengujian dan analisis performa model transformer berbasis IndoBERT dan IndoBERTweet dalam klasifikasi emosi pada unggahan media sosial X terkait GERD. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta didukung oleh visualisasi *confusion matrix*, grafik perbandingan performa, dan tabel ringkasan hasil evaluasi.

3.1 Hasil Pengumpulan Data

Data penelitian diperoleh melalui proses web crawling pada media sosial X menggunakan kata kunci yang berkaitan dengan *Gastroesophageal Reflux Disease* (GERD). Proses tersebut menghasilkan 1.440 tweet berbahasa Indonesia

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

yang berisi pengalaman, keluhan, dan opini pengguna terkait GERD. Seluruh data hasil crawling disimpan dalam format *Comma Separated Values* (CSV). Jumlah dataset relatif terbatas karena penelitian hanya menggunakan unggahan yang relevan, berbahasa Indonesia, tidak duplikat, dan memiliki konteks emosi yang dapat diidentifikasi. Data yang kosong, tidak relevan, bersifat promosi, atau tidak menunjukkan konteks emosi dieliminasi agar kualitas dataset tetap terjaga.

3.2 Data Preprocessing

Tahap *preprocessing* dilakukan untuk membersihkan dan menyiapkan data teks agar dapat digunakan secara optimal dalam proses analisis [22]. Data hasil *crawling* umumnya masih mengandung berbagai *noise*, seperti URL, emoji, tanda baca, serta penggunaan bahasa tidak baku yang dapat mempengaruhi kualitas analisis. Oleh karena itu, dilakukan beberapa tahapan *preprocessing*, meliputi *cleaning* untuk menghapus elemen yang tidak relevan, *case folding* untuk mengubah seluruh teks menjadi huruf kecil, serta normalisasi kata menggunakan kamus kata tidak baku agar teks menjadi lebih standar dan konsisten. Hasil dari tahapan ini berupa data teks yang lebih bersih dan siap digunakan pada proses pelabelan emosi.

	created_at	full_text	clean_text
0	Fri Oct 31 20:40:57 +0000 2025	Ahh gerd sialan	ahh gerd sialan
1	Fri Oct 31 20:28:13 +0000 2025	bisa nya ada org ovtin gue sampe gerd dia kamb...	bisa nya ada org ovtin gue sampe gerd dia kamb...
2	Fri Oct 31 19:58:26 +0000 2025	> anak2 CF banyak yg dari luar Jabodetabek ...	anak cf banyak yg dari luar jabodetabek masak ...
3	Fri Oct 31 18:58:02 +0000 2025	gerd https://t.co/zdEm9xn5bL	gerd
4	Fri Oct 31 18:52:01 +0000 2025	Sekalinya makan seblak abis itu panas sedada d...	sekalinya makan seblak abis itu panas sedada d...
5	Fri Oct 31 18:11:47 +0000 2025	PRO EM1 Original 90ml | Suplemen Probiotik | A...	pro em original ml suplemen probiotik anak dew...

Gambar 2. Hasil data preprocessing

3.3 Pelabelan Data

Tahap pelabelan data dilakukan untuk mengelompokkan setiap teks ke dalam kategori emosi tertentu, yaitu *angry*, *fear*, *joy*, *sad*, dan *netral*. Pelabelan dilakukan berdasarkan kemunculan kata atau konteks yang merepresentasikan emosi pada setiap teks. Hasil pelabelan menunjukkan adanya ketidakseimbangan distribusi kelas, dengan kelas *sad* sebagai kelas mayoritas sebanyak 436 data, diikuti *joy* sebanyak 280 data, *angry* sebanyak 262 data, *fear* sebanyak 219 data, dan *netral* sebanyak 146 data. Kondisi tersebut menjadi salah satu tantangan utama dalam proses klasifikasi emosi sehingga diperlukan penerapan teknik augmentasi pada tahap pelatihan model.

	created_at	full_text	clean_text	label
0	Fri Oct 31 20:40:57 +0000 2025	Ahh gerd sialan	ahh gerd sialan	angry
1	Fri Oct 31 20:28:13 +0000 2025	bisa nya ada org ovtin gue sampe gerd dia kamb...	bisa nya ada org ovtin gue sampe gerd dia kamb...	joy
2	Fri Oct 31 12:25:55 +0000 2025	@Ariaviersa waduh tapi kalo bawa makanan dari ...	waduh tapi kalo bawa makanan dari rumah boleh ...	sad
3	Fri Oct 31 12:14:47 +0000 2025	Hari ini sedih campur chaos bangeet aku hadapi...	hari ini sedih campur chaos bangeet aku hadapi...	fear
4	Fri Oct 31 12:02:21 +0000 2025	Sbg seseorang yg puny bejibun alergi dan gerd ...	sbg seseorang yg puny bejibun alergi dan gerd ...	angry
5	Fri Oct 31 11:33:29 +0000 2025	btw ini gerd emang karena ngopi WKWKWK	btw ini gerd emang karena ngopi wkwkwk	joy

Gambar 3. Hasil pelabelan emosi

Tabel 2. Jumlah data

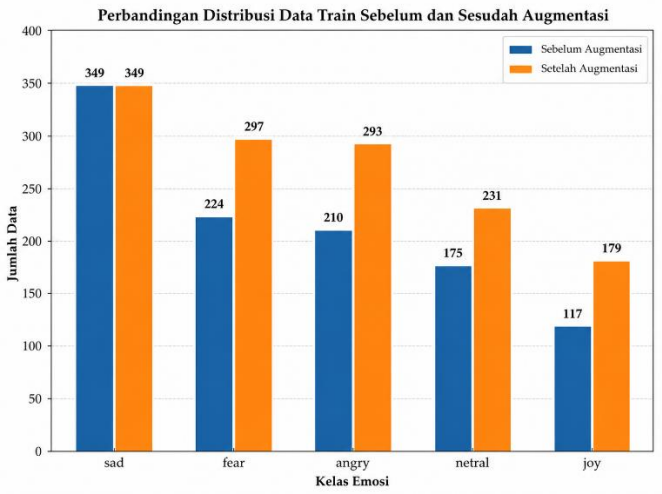
Label	Jumlah Data
Sad (Sedih)	436
Joy (Senang)	280
Angry (Marah)	262
Fear (Takut)	219
Netral	146

3.4 *Data Augmentation*

Untuk mengatasi ketidakseimbangan kelas, penelitian ini menerapkan *data augmentation* berbasis substitusi leksikal pada data pelatihan di setiap *fold* setelah proses *Stratified K-Fold Cross Validation*. Pendekatan ini bertujuan meningkatkan representasi kelas minoritas tanpa mengubah data asli maupun menyebabkan *data leakage*. Hasil augmentasi menunjukkan distribusi data pelatihan menjadi lebih seimbang sehingga model dapat mempelajari pola emosi secara lebih optimal.

Copyright © 2026 **Author**, Page 659
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039



Gambar 4. Hasil augmentasi

Hasil augmentasi menunjukkan peningkatan jumlah data pada kelas minoritas, seperti *joy*, *netral*, *angry*, dan *fear*, sementara kelas *sad* tetap menjadi kelas mayoritas. Perbedaan distribusi pada setiap *fold* dipengaruhi oleh proses *Stratified K-Fold Cross Validation*. Selain itu, data hasil augmentasi hanya diterapkan pada data pelatihan dan tidak digunakan pada data validasi maupun data uji untuk menjaga validitas evaluasi model.

3.5 Arsitektur Model

Arsitektur model dalam penelitian ini menggunakan dua pretrained language models berbasis Transformer, yaitu IndoBERT dan IndoBERTweet. Kedua model dipilih karena memiliki karakteristik data pelatihan yang berbeda. IndoBERT dilatih menggunakan korpus bahasa Indonesia umum, seperti artikel berita dan sumber ensiklopedia [23], sedangkan IndoBERTweet dilatih menggunakan data media sosial, khususnya Twitter [15]. Perbedaan ini digunakan untuk membandingkan kemampuan model dalam memahami bahasa formal dan informal pada klasifikasi emosi media sosial.

Kedua model menggunakan BertForSequenceClassification sebagai dasar klasifikasi. Teks diproses melalui tokenizer dengan panjang maksimum 128 token, lalu masuk ke encoder Transformer untuk menghasilkan representasi kontekstual. Representasi tersebut diteruskan ke classification head yang terdiri dari dropout 0,1 dan linear layer untuk memetakan teks ke lima kelas emosi, yaitu angry, fear, joy, netral, dan sad.

Pada full fine-tuning, seluruh parameter encoder dan classification head dilatih kembali. Pada LoRA, sebagian besar parameter pretrained model dibekukan, sedangkan parameter tambahan hanya dilatih pada modul attention, khususnya query dan value.

3.6 Fine-Tuning Konvensional

Penelitian ini menerapkan *full fine-tuning* sebagai pembanding terhadap pendekatan PEFT berbasis LoRA pada model IndoBERT dan IndoBERTweet. Pada *full fine-tuning*, seluruh parameter model diperbarui selama pelatihan, sedangkan pada LoRA hanya parameter tambahan tertentu yang dilatih. Kedua pendekatan menggunakan konfigurasi yang sama, yaitu *Stratified K-Fold Cross Validation, Focal Loss, dan early stopping*. Perbandingan difokuskan pada performa klasifikasi melalui accuracy, precision, recall, dan F1-score, serta perbedaan cakupan parameter yang diperbarui, bukan pada analisis biaya komputasi secara langsung.

3.7 Parameter-Efficient Fine-Tuning (LoRA)

Parameter-Efficient Fine-Tuning (PEFT) berbasis *Low-Rank Adaptation* (LoRA) digunakan sebagai pendekatan pelatihan yang hanya memperbarui sebagian kecil parameter tambahan tanpa melatih seluruh parameter model utama. Pendekatan ini dibandingkan dengan *full fine-tuning* pada model IndoBERT dan IndoBERTweet untuk melihat perbedaan performa klasifikasi dan cakupan parameter yang dilatih. Hasil eksperimen menunjukkan bahwa performa LoRA masih berada di bawah *full fine-tuning* pada metrik *accuracy*, *precision*, *recall*, dan F1-*score*. Namun, LoRA memiliki jumlah *trainable parameters* yang jauh lebih kecil dibandingkan *full fine-tuning*. Dengan demikian, pembahasan efisiensi dalam penelitian ini dibatasi pada jumlah parameter yang dilatih, bukan pada pengukuran waktu pelatihan atau penggunaan memori GPU.

3.8 Desain Eksperimen *(K-Fold Cross Validation)*

Penelitian ini menggunakan Stratified K-Fold Cross Validation sebagai metode pembagian data untuk menghasilkan evaluasi model yang lebih stabil dan representatif. Dataset dibagi ke dalam beberapa fold dengan tetap mempertahankan proporsi label pada setiap fold. Pada setiap iterasi, satu fold digunakan sebagai data uji, sedangkan

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

fold lainnya digunakan sebagai data latih. Pendekatan ini dipilih agar evaluasi tidak bergantung pada satu pembagian data saja, sehingga performa model dapat diamati secara lebih menyeluruh.

Epoch	Training Loss	Validation Loss
1	1.020973	1.003428
2	0.990665	0.992286
3	0.980729	0.989249

Accuracy Fold-1: 0.2876
F1 Macro Fold-1: 0.1551
F1 Weighted Fold-1: 0.1980

Epoch	Training Loss	Validation Loss
1	1.031034	0.992058
2	0.992547	0.978804
3	0.988863	0.975603

Accuracy Fold-5: 0.2685
F1 Macro Fold-5: 0.1424
F1 Weighted Fold-5: 0.1819

Gambar 5. Hasil fold pertama dan kelima pada IndoBERTweet

Berdasarkan hasil evaluasi Fold-1 dan Fold-5 pada model IndoBERTweet, training loss dan validation loss cenderung menurun pada setiap epoch. Pola ini menunjukkan bahwa proses pelatihan berjalan stabil. Namun, nilai accuracy dan F1-score macro pada beberapa fold masih rendah. Kondisi ini menunjukkan bahwa penurunan loss belum selalu diikuti oleh kemampuan klasifikasi yang baik pada seluruh kelas emosi. Performa yang rendah dapat dipengaruhi oleh beberapa faktor, yaitu jumlah data yang terbatas, distribusi kelas yang tidak seimbang, serta adanya kemiripan ekspresi antar emosi. Pada teks media sosial, pengguna sering menyampaikan emosi dengan bahasa informal, singkatan, konteks implisit, dan keluhan fisik yang tumpang tindih dengan kondisi emosional. Akibatnya, model lebih sulit membedakan kelas yang memiliki makna berdekatan, terutama pada kelas minoritas.

Epoch	Training Loss	Validation Loss
1	1.032681	0.987708
2	1.013972	0.979751
3	0.998023	0.977008

Accuracy Fold-1: 0.2876
F1 Macro Fold-1: 0.1416
F1 Weighted Fold-1: 0.1850

Epoch	Training Loss	Validation Loss
1	1.012362	0.989964
2	0.996290	0.978366
3	0.983710	0.975004

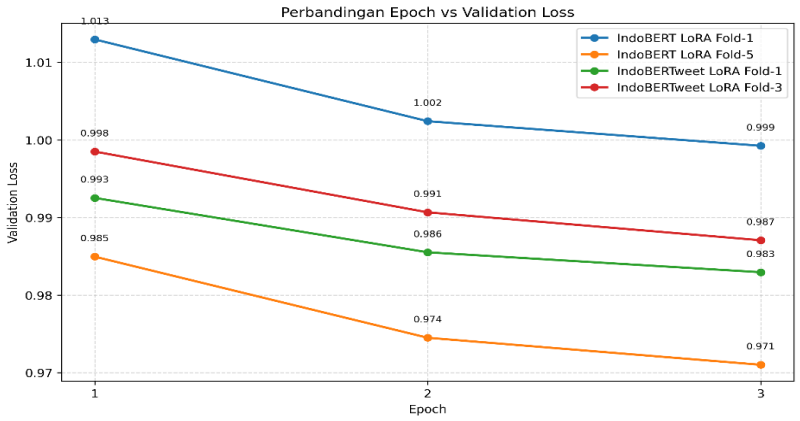
Accuracy Fold-5: 0.2987
F1 Macro Fold-5: 0.1683
F1 Weighted Fold-5: 0.2054

Gambar 6. Perbandingan fold pertama dan kelima pada IndoBERT

Hasil serupa juga terlihat pada model IndoBERT. Training loss dan validation loss menunjukkan pola penurunan, tetapi performa klasifikasi pada beberapa fold masih belum optimal. Hal ini mengindikasikan bahwa model mampu mempelajari pola umum dari data latih, namun belum sepenuhnya mampu melakukan generalisasi terhadap data uji. Nilai performa yang rendah pada beberapa fold dapat terjadi karena variasi komposisi data, keterbatasan jumlah sampel pada kelas minoritas, serta karakteristik teks GERD yang kompleks. Beberapa unggahan tidak hanya memuat ekspresi emosi, tetapi juga gejala fisik, keluhan kesehatan, dan pengalaman personal yang dapat memiliki makna emosional ganda. Oleh karena itu, hasil K-Fold menunjukkan bahwa klasifikasi emosi pada teks media sosial terkait GERD masih menjadi tugas yang menantang, khususnya dalam membedakan emosi dengan konteks yang berdekatan.

3.9 Early Stopping

Early stopping digunakan untuk mencegah *overfitting* dengan memantau nilai *validation loss* pada setiap *epoch* [24]. Proses pelatihan dihentikan secara otomatis apabila performa validasi tidak menunjukkan peningkatan dalam beberapa iterasi berturut-turut. Pendekatan ini bertujuan untuk menjaga stabilitas pelatihan dan mempertahankan kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.



Gambar 7. Perbandingan epoch vs validation loss

Copyright © 2026 **Author**, Page 661
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

3.10 Fungsi Loss *(Focal Loss)*

Focal Loss digunakan sebagai fungsi *loss* utama untuk membantu menangani ketidakseimbangan distribusi kelas pada dataset emosi GERD. Fungsi *loss* ini memberikan perhatian lebih besar pada sampel yang sulit diklasifikasikan, sehingga model tidak hanya berfokus pada pola kelas mayoritas. Dalam implementasinya, bobot alpha dihitung berdasarkan distribusi jumlah data pada masing-masing kelas di setiap *fold* pelatihan. Pendekatan ini digunakan untuk meningkatkan perhatian model terhadap kelas minoritas seperti *joy* dan netral. Penelitian ini tidak melakukan perbandingan langsung dengan *Cross-Entropy Loss,* sehingga pembahasan *Focal Loss* dibatasi pada perannya sebagai strategi penanganan class imbalance dalam proses pelatihan model.

3.11 *Hyperparameter*

Penentuan nilai hyperparameter dilakukan melalui manual tuning terbatas dengan mengacu pada konfigurasi yang umum digunakan dalam fine-tuning model BERT dan LoRA. Learning rate 2e-5 dipilih karena berada dalam rentang yang sering digunakan pada fine-tuning BERT, sedangkan batch size 8 disesuaikan dengan kapasitas GPU yang tersedia. Jumlah epoch ditetapkan sebanyak 5 dan dikombinasikan dengan early stopping untuk menjaga stabilitas pelatihan serta mengurangi risiko overfitting. Max length 128 digunakan karena data penelitian berupa teks pendek dari media sosial. Optimizer AdamW dan weight decay digunakan untuk membantu proses pembaruan parameter agar lebih stabil. Pada pendekatan LoRA, rank 8 dan alpha 16 digunakan sebagai konfigurasi adaptasi low-rank yang umum diterapkan untuk melatih parameter tambahan secara terbatas. Dengan demikian, strategi tuning pada penelitian ini bersifat manual berbasis referensi, bukan menggunakan grid search atau random search.

Tabel 3. Tabel Hyperparameter

Parameter	Nilai
Learning Rate	2e-5
Batch Size	8
Optimizer	AdamW
Epoch	5
Max Length	128
Weight Decay	0,05
LoRA Rank	8
LoRA Alpha	16
Dropout	0.1
Seed	42

3.12 Evaluasi Model

Evaluasi model dilakukan menggunakan *Stratified 5-Fold Cross Validation* untuk mengukur performa klasifikasi emosi pada teks media sosial terkait GERD menggunakan pendekatan *full fine-tuning* dan PEFT berbasis LoRA. Metrik evaluasi yang digunakan meliputi *accuracy*, *precision macro*, *recall macro*, *F1-score macro*, dan *F1-score weighted*, dengan *F1-score macro* sebagai metrik utama karena lebih representatif pada dataset tidak seimbang. Selain itu, nilai rata-rata dan standar deviasi antar *fold* dihitung untuk mengukur stabilitas performa model.

Hasil eksperimen menunjukkan bahwa pendekatan *full fine-tuning* secara konsisten menghasilkan performa lebih tinggi dibandingkan LoRA. IndoBERTweet dengan *full fine-tuning* memperoleh performa terbaik dengan nilai *accuracy* sebesar 0,5703, *precision macro* 0,5726, *recall macro* 0,5601, dan *F1-score macro* 0,5629. Sementara itu, IndoBERT memperoleh nilai *accuracy* sebesar 0,5145 dan *F1-score macro* sebesar 0,5034. Sebaliknya, model berbasis LoRA menunjukkan performa yang lebih rendah, yaitu *F1-score macro* sebesar 0,2412 pada IndoBERT dan 0,2227 pada IndoBERTweet. Hasil tersebut menunjukkan bahwa *full fine-tuning* lebih efektif dalam memahami kompleksitas semantik pada teks media sosial dibandingkan adaptasi parsial melalui LoRA.

Meskipun demikian, nilai standar deviasi yang relatif kecil pada seluruh model menunjukkan bahwa proses pelatihan berjalan cukup stabil antar *fold* dan tidak ditemukan indikasi *overfitting* yang signifikan selama proses pelatihan model.

Tabel 4. Hasil Perbandingan Full Fine-Tuning dan LoRA

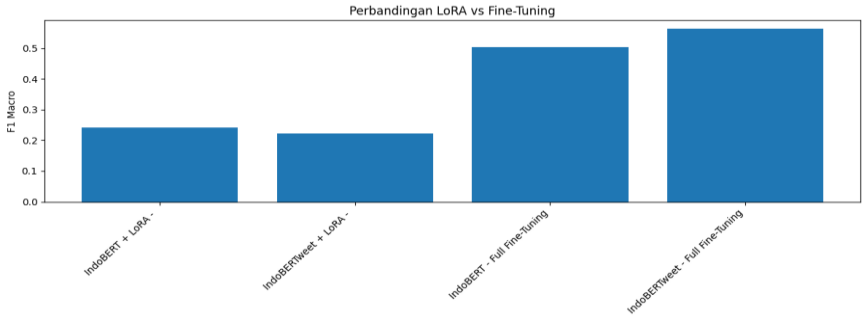
Model	Accuracy Mean	Accuracy Std	Precision Macro Mean	Precision Macro Std	Recall Macro Mean	Recall Macro Std	F1-Score Macro
IndoBERT	0,5145	±0,0456	0,5177	±0,0499	0,4993	±0,0391	0,5034
IndoBERTweet	0,5703	±0,0265	0,5726	±0,0304	0,5601	±0,0287	0,5629
IndoBERT + LoRA	0,2524	±0,0258	0,2528	±0,0321	0,2648	±0,0301	0,2412
IndoBERTweet + LoRA	0,2226	±0,0247	0,2373	±0,0191	0,2495	±0,0346	0,2227

Berdasarkan hasil evaluasi pada Tabel 4, pendekatan *full fine-tuning* menunjukkan performa yang lebih tinggi dibandingkan model berbasis LoRA pada seluruh metrik evaluasi. IndoBERTweet dengan *full fine-tuning*

Copyright © 2026 **Author**, Page 662
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

memperoleh performa terbaik dengan nilai *accuracy* sebesar 0,5703 dan *F1-score macro* sebesar 0,5629, sedangkan model berbasis LoRA hanya memperoleh *F1-score macro* sebesar 0,2412 pada IndoBERT dan 0,2227 pada IndoBERTweet. Hasil ini menunjukkan bahwa pembaruan seluruh parameter model masih lebih efektif dalam memahami kompleksitas semantik dan variasi emosi pada teks media sosial. Nilai standar deviasi yang relatif kecil juga menunjukkan bahwa performa model cenderung stabil antar *fold* tanpa indikasi *overfitting* yang signifikan.



Gambar 8. Perbandingan Performa LoRA dan Full Fine-Tuning

Visualisasi pada Gambar 8 menunjukkan bahwa model dengan *full fine-tuning* secara konsisten menghasilkan performa yang lebih baik dibandingkan LoRA, khususnya pada metrik *F1-score macro*. Meskipun demikian, LoRA tetap menunjukkan keunggulan dari sisi efisiensi komputasi karena hanya melatih sebagian kecil parameter model, sehingga menunjukkan adanya *trade-off* antara performa klasifikasi dan efisiensi sumber daya komputasi.

Tabel 5. Parameter efisiensi LoRA

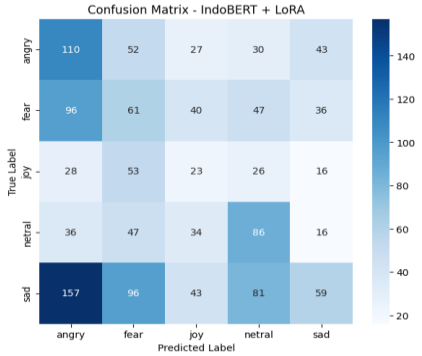
Model	Total Parameter	Trainable Parameter	Presentase Trainable
IndoBERT	124.445.189	124.445.189	100%
IndoBERTweet	110.562.309	110.562.309	100%
IndoBERT_LoRA	124.743.946	298.757	0,239%
IndoBERTweet_LoRA	110.860.810	298.757	0,269%

Berdasarkan Tabel 5, pendekatan LoRA menunjukkan efisiensi parameter yang signifikan dibandingkan *full fine-tuning*. Pada *full fine-tuning*, seluruh parameter model diperbarui selama pelatihan, sedangkan LoRA hanya melatih kurang dari 1% total parameter model. Meskipun lebih efisien dari sisi memori dan komputasi, performa klasifikasi LoRA masih berada di bawah *full fine-tuning*, yang menunjukkan keterbatasan dalam menangkap representasi emosional pada teks media sosial yang kompleks dan tidak terstruktur.

	Model	Accuracy Mean	Accuracy Std	Precision Macro Mean	Precision Macro Std	Recall Macro Mean	Recall Macro Std	F1 Macro Mean
0	IndoBERT + LoRA	0.252400	0.025854	0.252765	0.032133	0.264828	0.030069	0.241218
1	IndoBERTweet + LoRA	0.222638	0.024769	0.237296	0.019194	0.249539	0.034601	0.222670
2	IndoBERT	0.514479	0.045616	0.517731	0.049902	0.499395	0.039162	0.503430
3	IndoBERTweet	0.570349	0.026564	0.572629	0.030480	0.560087	0.028759	0.562924

Gambar 9. Perbandingan Performa LoRA dan Full Fine-Tuning

Untuk memperoleh pemahaman yang lebih mendalam mengenai distribusi kesalahan prediksi, dilakukan analisis menggunakan *confusion matrix* pada setiap pendekatan *fine-tuning*. Berdasarkan visualisasi pada Gambar 10 dan Gambar 10, model berbasis LoRA masih menunjukkan distribusi prediksi yang belum merata antar kelas. Pada IndoBERT + LoRA, sejumlah data dari kelas *fear*, *joy*, dan *sad* masih sering diprediksi sebagai *angry* atau *fear*, sedangkan IndoBERTweet + LoRA menunjukkan distribusi prediksi yang lebih menyebar namun masih mengalami kesulitan dalam membedakan emosi dengan kedekatan konteks.

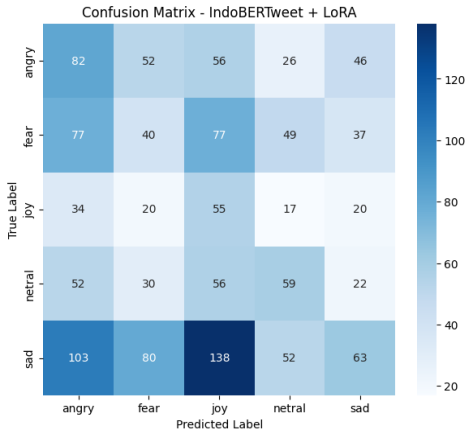


Gambar 10. Confusion Matrix IndoBERT + LoRA

Copyright © 2026 **Author**, Page 663
This Journal is licensed under a Creative Commons Attribution 4.0 International License

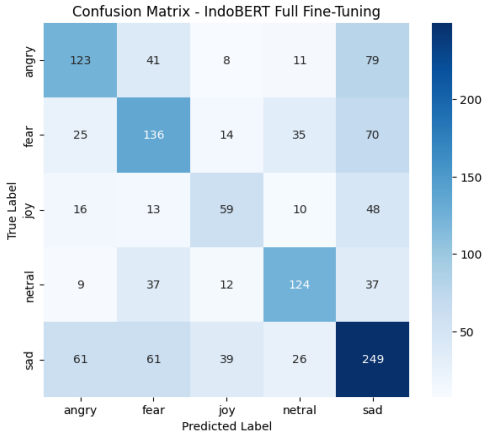
Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

Pada IndoBERT + LoRA, kelas *sad* menunjukkan tingkat kesalahan prediksi yang cukup tinggi karena sebagian besar data masih tersebar pada kelas *angry*, *fear*, dan *netral*. Selain itu, kelas *joy* juga masih sulit dikenali karena ekspresi emosi positif pada media sosial cenderung bersifat implisit.



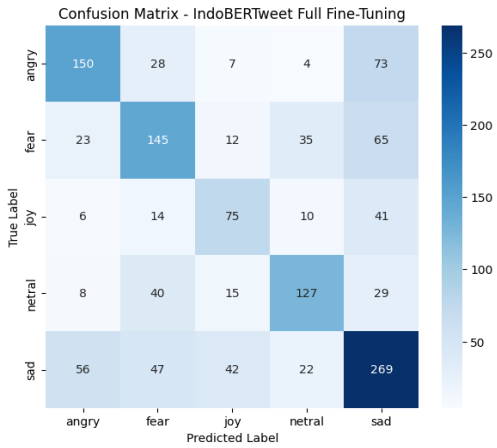
Gambar 11. Confusion Matrix IndoBERTweet + LoRA

Berbeda dengan pendekatan LoRA, model dengan *full fine-tuning* menunjukkan distribusi prediksi yang lebih terpusat pada diagonal *confusion matrix*, yang mengindikasikan peningkatan kemampuan model dalam mengenali kelas emosi yang sesuai. Pada IndoBERT dan IndoBERTweet *full fine-tuning*, kelas *fear*, *netral*, dan *sad* mulai dapat dikenali dengan lebih baik dibandingkan model berbasis LoRA.



Gambar 12. Confusion Matrix IndoBERT Full Fine-Tuning

Pada IndoBERTweet *full fine-tuning*, kemampuan model dalam mengenali kelas *angry*, *fear*, *netral*, dan *sad* menunjukkan peningkatan yang lebih baik dibandingkan model lainnya. Meskipun demikian, kelas *joy* masih menjadi salah satu kelas yang paling sulit dikenali secara konsisten karena ekspresi emosi positif pada teks media sosial cenderung ambigu dan kontekstual.



Gambar 13. Confusion Matrix IndoBERTweet Full Fine-Tuning

Copyright © 2026 **Author**, Page 664
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

Secara umum, model dengan *full fine-tuning* menunjukkan distribusi prediksi yang lebih terpusat pada diagonal *confusion matrix* dibandingkan model berbasis LoRA. Kondisi tersebut mengindikasikan bahwa *full fine-tuning* memiliki kemampuan yang lebih baik dalam mengenali hubungan antar kelas emosi secara konsisten. Sebaliknya, model berbasis LoRA masih mengalami kesulitan dalam membedakan beberapa kelas emosi yang memiliki kedekatan konteks, khususnya *sad*, *fear*, dan *angry*. Secara keseluruhan, hasil evaluasi menunjukkan bahwa klasifikasi emosi pada teks media sosial terkait GERD masih memiliki tingkat kompleksitas yang tinggi akibat penggunaan bahasa informal, ambiguitas ekspresi emosional, ketidakseimbangan distribusi kelas, serta tumpang tindih antar emosi.

Contoh error prediction menunjukkan bahwa model masih sulit membedakan emosi yang memiliki konteks berdekatan. Misalnya, teks "ternyata aku sakit gerd" berlabel asli netral, tetapi diprediksi sebagai sad karena mengandung kata "sakit" yang bernuansa negatif. Teks "yg gampang ngantuk kayanya gerd kak" juga berlabel asli fear, tetapi diprediksi sebagai sad karena memuat keluhan fisik yang dapat ditafsirkan sebagai ketidaknyamanan. Hal ini menunjukkan bahwa ekspresi emosi pada media sosial sering bersifat implisit, informal, dan memiliki makna ganda.

3.13 *Error Analysis*

Hasil *error analysis* menunjukkan bahwa model IndoBERT dan IndoBERTweet masih mengalami kesulitan dalam membedakan emosi yang memiliki kedekatan konteks, khususnya *fear*, *sad*, dan *netral*. Model juga cenderung menggeneralisasi berbagai emosi negatif ke dalam kelas *sad*. Kesalahan prediksi banyak ditemukan pada teks yang mengandung *mixed emotions*, humor, atau sarkasme, sedangkan kelas *joy* sering salah diklasifikasikan karena ekspresi emosi positif pada media sosial cenderung implisit. Selain itu, penggunaan bahasa informal, singkatan, serta campuran bahasa Indonesia dan bahasa asing turut mempengaruhi performa model. Secara keseluruhan, hasil *error analysis* menunjukkan bahwa kesalahan prediksi dipengaruhi oleh ketidakseimbangan kelas, ambiguitas makna, dan tingginya variasi bahasa pada media sosial, sehingga diperlukan peningkatan kualitas pelabelan dan pendekatan model yang lebih adaptif untuk penelitian selanjutnya.

3.14 Interpretasi Medis

Hasil klasifikasi emosi menunjukkan bahwa pembahasan GERD pada media sosial didominasi oleh emosi negatif, khususnya *sad*, yang mengindikasikan adanya keterkaitan antara kondisi fisiologis dan dimensi emosional pengguna. Kemunculan emosi *fear* yang sering tertukar dengan *sad* menunjukkan adanya kedekatan konteks antara kekhawatiran dan ketidaknyamanan akibat kondisi kesehatan yang dialami. Selain itu, kesulitan model dalam membedakan kelas *angry*, *fear*, dan *sad* menunjukkan bahwa ekspresi emosi pada penderita GERD bersifat kompleks dan saling tumpang tindih. Sebaliknya, kelas *joy* relatif jarang muncul dan sulit dikenali karena ekspresi emosi positif cenderung disampaikan secara implisit pada media sosial. Secara keseluruhan, hasil ini menunjukkan bahwa pengalaman penderita GERD berpotensi melibatkan aspek emosional selain kondisi fisik. Namun, interpretasi dalam penelitian ini masih bersifat eksploratif dan tidak dimaksudkan sebagai diagnosis klinis, melainkan sebagai pendekatan awal berbasis NLP untuk memahami kecenderungan emosional pengguna melalui teks media sosial.

4. KESIMPULAN

Penelitian ini membandingkan performa IndoBERT dan IndoBERTweet dalam klasifikasi emosi penderita GERD pada media sosial menggunakan full fine-tuning dan PEFT berbasis LoRA. Hasil eksperimen menunjukkan bahwa full fine-tuning menghasilkan performa terbaik, dengan F1-score macro sebesar 0,5629 pada IndoBERTweet dan 0,5034 pada IndoBERT. Sementara itu, LoRA memperoleh F1-score macro sebesar 0,2412 pada IndoBERT dan 0,2227 pada IndoBERTweet. Temuan ini menunjukkan bahwa full fine-tuning lebih sesuai untuk dataset emosi GERD yang berukuran kecil, tidak seimbang, dan memiliki karakteristik bahasa informal. Meskipun performa LoRA lebih rendah, hasil ini tetap memberikan temuan penting bahwa adaptasi parameter terbatas belum cukup untuk menangkap kompleksitas ekspresi emosi pada teks kesehatan media sosial. LoRA tetap memiliki keunggulan dari sisi jumlah parameter yang dilatih karena hanya memperbarui kurang dari 1% total parameter model. Namun, efisiensi parameter tersebut belum menghasilkan performa klasifikasi yang sebanding dengan full fine-tuning pada konteks penelitian ini. Dengan demikian, kontribusi utama penelitian ini terletak pada pembuktian empiris bahwa pemilihan strategi fine-tuning perlu disesuaikan dengan ukuran dataset, distribusi kelas, dan kompleksitas domain teks. Hasil confusion matrix dan error analysis menunjukkan bahwa model masih mengalami kesulitan dalam membedakan emosi yang memiliki kedekatan konteks, terutama fear dan sad. Kesalahan prediksi banyak dipengaruhi oleh ketidakseimbangan kelas, penggunaan bahasa informal, serta tumpang tindih antara keluhan fisik GERD dan ekspresi emosional. Penelitian selanjutnya dapat meningkatkan performa model dengan menambah data berlabel pada kelas minoritas, memperbaiki konsistensi anotasi, menerapkan augmentasi kontekstual, serta menambahkan pengukuran training time dan GPU memory usage agar analisis efisiensi LoRA dapat diuji secara lebih lengkap.

REFERENCES

[1] M. S. Maqbul *et al.*, "Gastro-esophageal reflux disease among the urban population of Saudi Arabia," *Gastroenterology and Endoscopy*, vol. 2, no. 3, pp. 121–130, Jul. 2024, doi: 10.1016/j.gande.2024.07.004.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 655−666
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10039

[2] R. Syifa, A. Nirwani, Z. Arifin, and T. A. Laariya, "Hubungan Tingkat Stres Akademik dan Kejadian Gastroesophageal Reflux Disease (GERD) pada Mahasiswa S1 Program Studi Kedokteran, Fakultas Kedokteran Universitas Ahmad Dahlan." Accessed: Jun. 13, 2026. [Online]. Available: https://doi.org/10.55175/cdk.v52i8.1594

[3] L. Chen *et al.*, "Global, regional and national burdens of gastroesophageal reflux disease from 1990 to 2021 and projections to 2050: a finding from the global burden of disease study 2021," *BMC Public Health*, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12889-025-25121-w.

[4] S. Deasy Siregar, V. Trismanjaya Hulu, and D. Veronika Lase, "Factors Associated With The Occurrence Of Gastroesophageal Reflux Disease (Gerd) In Students," *Jurnal Berkala Epidemiologi*, vol. 13, no. 1, pp. 58–65, 2025, doi: 10.20473/jbe.v13i12025.58.

[5] S. Murillo, "The use of emojis in X/Twitter for research recontextualization," *Discourse, Context and Media*, vol. 68, Dec. 2025, doi: 10.1016/j.dcm.2025.100950.

[6] S. Hinduja, M. Afrin, S. Mistry, and A. Krishna, "Machine learning-based proactive social-sensor service for mental health monitoring using twitter data," *International Journal of Information Management Data Insights*, vol. 2, no. 2, Nov. 2022, doi: 10.1016/j.jjimei.2022.100113.

[7] E. Mayor, M. Miché, and R. Lieb, "Associations between emotions expressed in internet news and subsequent emotional content on twitter," *Heliyon*, vol. 8, no. 12, Dec. 2022, doi: 10.1016/j.heliyon.2022.e12133.

[8] W. Subait *et al.*, "Artificial Intelligence-based Natural Language Processing for sarcasm detection and classification on Arabic Corpus," *Alexandria Engineering Journal*, vol. 125, pp. 320–331, Jun. 2025, doi: 10.1016/j.aej.2025.03.125.

[9] J. I. T. Krisna, A. Luthfiarta, L. D. Cahya, S. Winarno, and A. Nugraha, "Comparing Optimizer Strategies For Enhancing Emotion Classification In IndoBERT Models," *Advance Sustainable Science, Engineering and Technology*, vol. 6, no. 2, Feb. 2024, doi: 10.26877/asset.v6i2.18228.

[10] A. Amelia and R. Yusuf, "Analisis Sentimen Masyarakat Indonesia Pada Platfrom X Terhadap Isu Fufufafa Menggunakan Bidirectional Encoder Representations From Transformers," 2025. Accessed: Jun. 13, 2026. [Online]. Available: 10.51401/jinteks.v7i1.5160

[11] R. Merdiansah 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, 2024, Accessed: Jun. 13, 2026. [Online]. Available: ejournal.sisfokomtek.org/index.php/jikom

[12] N. M. Damayanti, "Analisis Sentimen Publik Pada Tagar #Btscomeback Di Platform X Menggunakan Indobertweet," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, Jul. 2025, doi: 10.23960/jitet.v13i3.7176.

[13] D. Ismiyana Putri, A. Nurul Alfian, M. Yudhi Putra, and P. Dwi Mulyo, "IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election," 2024. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

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

[15] U. Khairani, V. Mutiawani, and H. Ahmadian, "Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 2024.

[16] D. Nur Wahyu Ningsih, "Deteksi Risiko Depresi Pada Pengguna Media Sosial Indonesia Menggunakan Indobertweet," 2025. Accessed: Jun. 13, 2026. [Online]. Available: journal.risetdigital.org/JCSC/article/view/8/3

[17] N. Marito Putry and B. Nurina Sari, "Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus," *Jurnal Sains dan Manajemen*, vol. 10, no. 1, 2022, Accessed: Jun. 13, 2026. [Online]. Available: doi.org/10.31294/EVOLUSI.V10I1.12514

[18] I. Topan Adib Amrulloh *et al.*, "Evaluasi Augmentasi Data Pada Deteksi Penyakit Daun Tebu Dengan YOLOV8," 2024. Accessed: Jun. 13, 2026. [Online]. Available: 10.36040/jati.v8i4.10267

[19] R. Fadhilah, F. Hakim, C. V. Pangemanan, E. Supriyadi, and U. Koperasi Indonesia, "CoopMind V1: Fine-Tuning Large Language Model dengan Pendekatan Low Rank Adaptation (LoRA) untuk Demokratisasi Pengetahuan Perkoperasian," 2026. Accessed: Jun. 13, 2026. [Online]. Available: journal2.ikopin.ac.id/index.php/dataenthusiast/article/view/17/19

[20] K. Davis and R. Maiden, "The Importance of Understanding False Discoveries and the Accuracy Paradox When Evaluating Quantitative Studies," *Stud. Soc. Sci. Res.*, vol. 2, no. 2, p. p1, Mar. 2021, doi: 10.22158/sssr.v2n2p1.

[21] F. Zen *et al.*, "Analisis Perbandingan Efektivitas Arsitektur Convolutional Neural Network Dengan Focal Loss Pada Kasus Imbalance Data Klasifikasi Hewan," 2025, doi: 10.32485/jcsai.

[22] F. Reynaldi Valerian, M. Syarief, D. Abdul Fatah, J. Raya Telang, K. Kamal, and J. Timur, "Klasifikasi Tingkat Obesitas Menggunakan Metode Gbm Dan Confusion Matrix," 2025. Accessed: Jun. 13, 2026. [Online]. Available: doi.org/10.36040/jati.v9i2.13062

[23] A. Gumelar Syah Moeslim, Esa Firmansyah, and Beben 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.

[24] M. Z. Haq, C. S. Octiva, A. Ayuliana, U. W. Nuryanto, and D. Suryadi, "Algoritma Naïve Bayes untuk Mengidentifikasi Hoaks di Media Sosial," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 1079–1084, Jul. 2024, doi: 10.33395/jmp.v13i1.13937.

This Journal is licensed under a Creative Commons Attribution 4.0 International License