

Optimasi Pengenalan Emosi dalam Ulasan Produk Berbahasa Indonesia dengan LoRA pada Transformer

Optimizing Emotion Recognition in Indonesian Product Reviews Using LoRA on Transformer

David Suharjanto^{1*}, Muhammad Mustakim²

^{1,2} Informatika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga, Yogyakarta, Indonesia

^{1*} davidsuharjanto7@gmail.com, ² m.mustaqim@uin-suka.ac.id

Abstract

Emotion analysis in product reviews plays a crucial role in Indonesia's e-commerce landscape, yet training advanced transformer models demands high computational resources, posing a significant constraint for small businesses. This research evaluates the Low-Rank Adaptation (LoRA) technique for optimizing emotion recognition in Indonesian product reviews, with a focus on resource efficiency. Indobert-large-p2, Indobert-base-uncased, and Indobertweet-base-uncased models were trained on the PRDECT-ID dataset, comparing the performance of full fine-tuning and LoRA. Results show LoRA provided competitive performance; on the Indobert-large-p2 model, the F1-score reached 67.81%, surpassing full fine-tuning's 67.74%, despite training only about 1% of the total parameters. LoRA significantly reduced VRAM consumption by up to 17% and accelerated training duration by up to four minutes. Furthermore, the transformer models fine-tuned in this study, using both LoRA and full fine-tuning, consistently outperformed the results of previous research that employed complex architectures like CNN and BiLSTM, with the highest F1-score reaching 69.72% compared to the previous best of 66.13%. LoRA proves to be a practical and efficient solution for limited computational resources, enabling effective emotion analysis without expensive infrastructure.

Keywords: Emotion Recognition; LoRA; Transformer; Product Reviews; Bahasa Indonesia

Abstrak

Analisis emosi pada ulasan produk memainkan peran penting dalam lanskap *e-commerce* di Indonesia, namun pelatihan model transformer yang canggih memerlukan sumber daya komputasi tinggi dapat menjadi kendala bagi perusahaan kecil. Penelitian ini mengevaluasi teknik *Low-Rank Adaptation* (LoRA) untuk optimasi pengenalan emosi pada ulasan produk berbahasa Indonesia, dengan fokus pada efisiensi sumber daya. Model Indobert-large-p2, Indobert-base-uncased, serta Indobertweet-base-uncased dilatih pada dataset PRDECT-ID, dengan perbandingan performa *full fine-tuning* dan LoRA. Hasil menunjukkan LoRA memberikan performa kompetitif, terutama pada model IndoBERT-large-p2, F1-score mencapai 67,81%, mengungguli hasil dari *full fine-tuning* 67,74%, meski hanya melatih sekitar 1% dari total parameter. LoRA secara signifikan mengurangi konsumsi VRAM hingga 17% dan mempercepat durasi pelatihan hingga empat menit. Lebih lanjut, model transformer yang di-*fine-tune* dalam studi ini, baik dengan LoRA maupun *full fine-tuning*, secara konsisten mengungguli hasil dari penelitian sebelumnya yang menggunakan tambahan arsitektur kompleks seperti CNN dan BiLSTM, dengan F1-score tertinggi mencapai 69,72% dibandingkan hasil terbaik pada penelitian sebelumnya (66,13%). LoRA terbukti menjadi solusi praktis dan efisien bagi sumber daya komputasi terbatas, memungkinkan analisis emosi efektif tanpa infrastruktur mahal.

Kata kunci: Pengenalan Emosi; LoRA; Transformer; Ulasan Produk; Bahasa Indonesia

1. Pendahuluan

Dalam beberapa tahun terakhir, pertumbuhan pesat *e-commerce* di Indonesia telah mengubah lanskap perdagangan dan perilaku konsumen secara signifikan [1], [2]. Platform *e-commerce* seperti Shopee dan Tokopedia tidak hanya menjadi tempat jual beli, tetapi juga wadah interaksi antara pelanggan dan penjual melalui ulasan produk [3]. Ulasan ini berperan sebagai sumber informasi yang sangat penting dalam proses pengambilan keputusan konsumen, karena

memberikan gambaran langsung dari pengalaman pengguna sebelumnya [4]. Tidak hanya menyampaikan penilaian terhadap kualitas produk, ulasan juga mencerminkan persepsi terhadap layanan, kecepatan pengiriman, dan bahkan kepuasan dari pelanggan [5]. Selain itu, ulasan pelanggan juga merupakan bentuk *electronic word-of-mouth* (e-WOM) yang memiliki daya pengaruh tinggi terhadap calon pembeli [6]. Dengan demikian, ulasan positif mampu meningkatkan kepercayaan pelanggan

terhadap merek dan memperbesar peluang terjadinya pembelian, sedangkan ulasan yang negatif dapat menurunkan persepsi kualitas suatu produk [7], [8], [9].

Lebih jauh, ulasan produk juga menyediakan wawasan yang berharga bagi pelaku usaha [10]. Dengan menganalisis isi ulasan, perusahaan dapat mengevaluasi performa produk mereka secara lebih objektif, mengidentifikasi kelemahan layanan, serta menyusun strategi peningkatan yang lebih tepat sasaran [11], [12]. Dalam skala besar, akumulasi ulasan pelanggan menjadi data tak ternilai yang mencerminkan hubungan antara merek dan konsumennya secara *real-time* [13]. Oleh karena itu, penting bagi perusahaan untuk tidak hanya memperhatikan keberadaan ulasan, tetapi juga memahami emosi yang terkandung di dalamnya, karena aspek emosional sering kali menjadi indikator tersembunyi dari loyalitas, kepuasan, atau bahkan potensi kehilangan pelanggan [14].

Namun demikian, menganalisis emosi dalam ulasan pelanggan secara manual merupakan proses yang melelahkan dan tidak efisien, terutama ketika data yang tersedia dalam jumlah besar [15]. Hal ini disebabkan karena emosi yang terkandung dalam ulasan sering kali tidak diungkapkan secara eksplisit, melainkan tersirat melalui penilaian positif atau negatif yang kemudian ditafsirkan ke dalam kategori emosi seperti marah, takut, senang, cinta, atau sedih [16]. Tantangan lain muncul dari penggunaan kalimat tidak baku, penggunaan emotikon, dan simbol non-verbal yang banyak ditemukan dalam ulasan produk [17]. Oleh karena itu, otomatisasi dalam pengenalan emosi menjadi kebutuhan mendesak, khususnya bagi perusahaan yang ingin memahami persepsi pelanggan secara cepat dan konsisten [18].

Berbagai upaya pendekatan telah dilakukan untuk otomatisasi pengenalan emosi dalam domain *Natural Language Processing* (NLP). Misalnya, Saputri et al. [19] mengembangkan sistem klasifikasi emosi berbasis fitur-fitur seperti *lexicon-based*, *Bag-of-Words*, *word embeddings*, *orthography*, dan *Part-of-Speech (POS) tags*. Dengan menggunakan dataset berisi 4.403 cuitan dari Twitter, model tersebut mencapai F1-score sebesar 69%. Sementara itu, Haryadi & Kusuma [20], menerapkan pendekatan berbasis *deep learning* yang mengimplementasikan *Long Short-Term Memory* (LSTM) dan *Nested LSTM* untuk mengklasifikasikan tujuh kategori emosi dari dataset berisi sekitar satu juta tweet. Hasilnya menunjukkan performa yang sangat tinggi, dengan akurasi dan F1-score masing-masing mencapai 99%. Penelitian lain oleh Lie et al. [21] mengadopsi arsitektur *Bi-directional Gated Recurrent Unit* (Bi-GRU) yang dipadukan dengan *Attention Mechanism* pada dataset NLPCCData2013 dan NLPCCData2014 yang terdiri dari 25 ribu data, menghasilkan F1-score sebesar 69%.

Meskipun sebagian besar penelitian tersebut belum secara langsung diterapkan pada konteks *e-commerce*, arah dan tujuannya tetap sejalan, yaitu mengenali emosi dalam teks tidak terstruktur secara otomatis. Umumnya, penelitian-penelitian terdahulu berfokus pada data dari media sosial seperti Twitter atau korpus internasional yang sudah banyak tersedia, namun belum banyak yang mengkaji pengenalan emosi dalam ulasan produk dari platform *e-commerce* Indonesia secara spesifik. Padahal, ulasan produk di *e-commerce* juga memuat ekspresi emosional pengguna yang penting bagi evaluasi layanan dan strategi bisnis. Salah satu penelitian yang secara khusus menggunakan dataset ulasan produk *e-commerce* berbahasa Indonesia yaitu Chowanda et al. [22], dengan memanfaatkan dataset PRDECT-ID [23]. Dataset ini terdiri dari 5.400 ulasan dari 29 produk Tokopedia, yang telah dilabeli ke dalam lima kategori emosi. Namun, dataset ini memiliki tantangan berupa ketidakseimbangan distribusi label, yang dapat memengaruhi performa model klasifikasi.

Dalam penelitian tersebut, model BERT dikombinasikan dengan CNN maupun LSTM. Hasil terbaik diperoleh dari kombinasi BERT dengan CNN, yang mencapai akurasi sebesar 69,26% dan F1-score sebesar 65,49% [22]. Hasil ini menunjukkan bahwa penggabungan arsitektur mampu meningkatkan performa, namun pada saat yang sama berpotensi menambah kompleksitas model dan kebutuhan komputasi. Kompleksitas ini dapat menjadi hambatan bagi perusahaan kecil dengan sumber daya terbatas, baik dari segi waktu pelatihan maupun kebutuhan VRAM pada GPU. Oleh karena itu, terdapat peluang untuk melakukan optimasi dengan pendekatan yang lebih efisien.

Seiring berkembangnya *pre-trained language model*, penelitian terkini mulai beralih dari arsitektur berbasis CNN dan RNN menuju model transformer seperti IndoBERT dan IndoRoBERTa yang mampu menghasilkan representasi kontekstual lebih baik untuk tugas klasifikasi emosi berbahasa Indonesia. Beberapa penelitian menunjukkan bahwa model transformer memberikan performa kompetitif pada berbagai tugas klasifikasi emosi [24], [25].

Selain itu, perkembangan terbaru dalam *Parameter-Efficient Fine-Tuning* (PEFT) telah menghasilkan berbagai pendekatan seperti Adapter, Prefix-Tuning, QLoRA, dan LoRA telah dikembangkan untuk mengurangi kebutuhan komputasi selama proses *fine-tuning* tanpa mengorbankan performa model [26]. Adapter menambahkan modul pelatihan baru pada setiap layer Transformer [27], sedangkan Prefix-Tuning mengoptimalkan representasi awalan (*prefix*) yang disisipkan pada mekanisme *self-attention* [28]. QLoRA (*Quantized Low-Rank Adaptation*) merupakan pengembangan dari LoRA (*Low-Rank Adaptation*) yang mengombinasikan kuantisasi model dengan pembaruan parameter berperingkat rendah (*low-rank*)

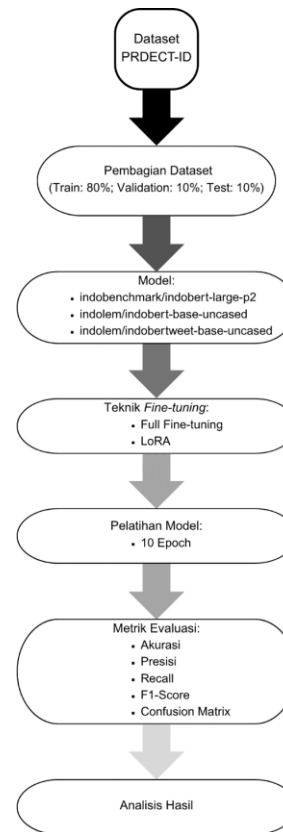
sehingga memungkinkan *fine-tuning* model *Large Language Models* dengan kebutuhan memori dan komputasi yang lebih rendah [29]. Pada penelitian ini, LoRA dipilih karena memiliki implementasi yang sederhana, kompatibel dengan model encoder seperti IndoBERT, serta mampu mengurangi jumlah parameter yang dilatih tanpa memerlukan kuantisasi maupun penambahan modul arsitektur yang kompleks.

Teknik *Low-Rank Adaptation* (LoRA) dapat menjadi solusi yang menjanjikan [30]. LoRA merupakan pendekatan PEFT yang memungkinkan adaptasi model besar seperti transformer hanya dengan menambahkan sejumlah kecil parameter tambahan [31]. Pendekatan ini tidak hanya mengurangi risiko *overfitting*, tetapi juga menghemat penggunaan sumber daya komputasi secara signifikan tanpa mengorbankan performa prediksi. Dengan mempertimbangkan celah penelitian tersebut, studi ini bertujuan untuk mengeksplorasi efektivitas penerapan LoRA pada model transformer *pre-trained* berbahasa Indonesia untuk tugas klasifikasi emosi dalam ulasan produk. Fokus utama penelitian ini adalah membandingkan performa dan efisiensi dari *fine-tuning* menggunakan LoRA dengan pendekatan *full fine-tuning*, serta juga akan membandingkan performa model dengan hasil yang diperoleh dari penelitian sebelumnya yang cenderung menggunakan arsitektur model yang lebih kompleks.

2. Metodologi Penelitian

Bagian ini menjelaskan secara rinci metodologi yang diterapkan dalam penelitian, meliputi desain eksperimen, arsitektur model, detail dataset, lingkungan pengembangan, serta konfigurasi pelatihan, dan metrik evaluasi. Keseluruhan proses penelitian digambarkan secara komprehensif seperti pada Gambar 1.

Proses ini dimulai dengan persiapan dataset PRDECT-ID [23], yang merupakan kumpulan ulasan produk berbahasa Indonesia. Dataset ini kemudian melalui tahap pembagian dataset, di mana data dibagi menjadi tiga bagian utama yaitu 60% untuk *training* (pelatihan), 20% untuk *validation* (validasi), dan 20% untuk *test* (pengujian). Setelah dataset siap, langkah selanjutnya adalah memilih dan menginisialisasi model transformer *pre-trained*. Dalam studi ini, tiga model transformer yaitu *indobert-large-p2*, *indobert-base-uncased*, dan *indoberttweet-base-uncased* digunakan sebagai *backbone*. Model-model ini kemudian dilatih melalui pelatihan model selama 10 epoch menggunakan dua teknik *fine-tuning* yang berbeda yaitu *full fine-tuning* dan LoRA. Setelah pelatihan selesai, kinerja model dievaluasi secara sistematis menggunakan metrik evaluasi standar seperti akurasi, presisi, recall, F1-Score, dan *confusion matrix*. Hasil dari evaluasi ini selanjutnya menjadi dasar untuk analisis hasil guna mendapatkan temuan dan kesimpulan penelitian.



Gambar 1. Diagram Alir Penelitian

2.1. Arsitektur Model

Penelitian ini memanfaatkan tiga model transformer *pre-trained* berbahasa Indonesia yang telah terbukti kemampuannya dalam memahami nuansa kompleks Bahasa Indonesia dan menunjukkan performa unggul dalam berbagai tugas *Natural Language Processing* (NLP). Model pertama dan kedua yaitu *Indobert-base-uncased* (124,5 juta parameter) dan *Indobert-large-p2* (335 juta parameter), yang diperkenalkan oleh Koto et al. [32]. *Indobert* merupakan adaptasi *Bidirectional Encoder Representations from Transformers* (BERT) khusus untuk Bahasa Indonesia, yang dilatih menggunakan korpus data yang beragam, termasuk Wikipedia, Kompas, hingga data dari korpus web Indonesia.

Sementara itu, model ketiga adalah *Indoberttweet-base-uncased* (110 juta parameter), yang dikembangkan oleh Koto et al. [33]. Model ini merupakan varian BERT yang dilatih secara spesifik pada data Twitter berbahasa Indonesia, sehingga sangat adaptif terhadap karakteristik bahasa informal dan singkat yang sering ditemukan dalam ulasan daring. Ketiga model ini dipilih karena potensinya yang besar untuk digunakan sebagai *backbone* dalam tugas pengenalan emosi pada teks berbahasa Indonesia, mengingat arsitektur dan proses pelatihannya yang disesuaikan dengan karakteristik yang sama.

2.2. Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset ulasan produk berbahasa Indonesia yang dikembangkan oleh Sutoyo et al. [23]. Dataset ini terdiri dari total 5.400 ulasan produk berbahasa Indonesia dari Tokopedia yang telah dianotasi, dan dikategorikan menjadi lima label yaitu Cinta, Takut, Sedih, Senang, dan Marah. Pembagian dataset dilakukan menggunakan *stratified random split* (*random seed* = 42) dengan proporsi 60% untuk *training set*, 20% untuk *validation set*, dan 20% untuk *test set*. Detail pembagian dataset dapat dilihat pada Tabel 1, dan beberapa contoh ulasan beserta label emosinya disajikan pada Tabel 2.

Tabel 1. Dataset Ulasan Produk Berbahasa Indonesia

Kategori	Train	Validation	Test
Cinta	477	169	163
Takut	544	166	210
Sedih	725	246	231
Senang	1.069	361	340
Marah	425	138	136
Total	3.240	1.080	1.080

Tabel 2. Contoh Ulasan Produk dan Kategori Emosinya

Kategori Emosi	Contoh
Cinta	lucu bgt warnanya. recommended harga murah tapi ngga murahan. next bakal order lagi Barang mantap anak saya suka, dah gitu pengiriman cepat banget hari ini pesan besok dah sampe
Takut	hati yang belanja disini, suka kirim. barang sembarangan baru lama lagi buat returnnya, toko tidak niat kek gitu, mau nipu, tanpa video unboxing pasti ngk bs retur jadi ingat selalu video di toko ini, penipuan tokonya kaca malah jadi buram. cairan ini baunya seperti porstex. hati hati dalam membeli
Sedih	Barang tidak di kirim sampai sekarang, respon penjual payah dan mengecewakan, bikin kapok, kalau barang nggak ada tolong uang ku di kembalikan semua mengecewakan
Senang	Cepat dan packing amanah bgt. kualitas susu sdh terjamin enak dan bermanfaat. thanks. barang diterima dengan baik, kualitas baik, pengiriman cepat, makasih
Marah	MAC HANCUR, PADAHAL SUDAH ADD BUBBEL. KRN KARDUS CUMAN NEMPEL DOANK DI MAC, JD KETINDIH2 PECAH kalo mau kirim baca note nya dong, masa tiga2 nya dikirim warna sama, trus fungsinya catatan buat apa

2.3. Lingkungan Pengembangan

Seluruh eksperimen dan pelatihan model dilaksanakan menggunakan platform Google Colaboratory.

Lingkungan ini menyediakan akses ke GPU T4 dengan kapasitas VRAM hingga 15 GB, dan diimplementasikan dengan *framework deep learning* PyTorch, PEFT, serta Hugging Face Transformers untuk manajemen model dan tokenisasi.

2.4. Konfigurasi Fine-tuning

Proses *fine-tuning* model dilakukan menggunakan dua teknik utama yaitu *full fine-tuning* dan *Low-Rank Adaptation* (LoRA). Untuk LoRA, parameter yang digunakan adalah $r = 8$, lora alpha = 16, serta lora dropout = 0,1. Nilai tersebut dipilih dengan mengacu pada konfigurasi yang umum digunakan pada penelitian sebelumnya [34]. Modul LoRA diterapkan pada *layer query*, *key*, *value*, dan *dense* pada mekanisme *self-attention* transformer. *Learning rate* LoRA diatur ke $2e-4$, sedangkan untuk *full fine-tuning* adalah $2e-5$. Perbedaan *learning rate* tersebut mengacu pada praktik yang umum digunakan dalam implementasi LoRA, di mana LoRA umumnya menggunakan *learning rate* yang lebih tinggi dibandingkan *full fine-tuning* karena hanya mengoptimalkan sebagian kecil parameter yang dapat dilatih [35]. Konfigurasi hyperparameter umum untuk kedua teknik ini disajikan pada Tabel 3, sedangkan detail konfigurasi spesifik dapat dilihat pada Tabel 4. Seluruh eksperimen menggunakan *random seed* = 42 dan dilakukan satu kali pelatihan (*single run*) untuk setiap konfigurasi model.

Tabel 3. Konfigurasi Umum Hyperparameter

Hyperparameter	Nilai
<i>Optimizer</i>	AdamW
<i>LR scheduler</i>	Cosine
<i>Training batch size</i>	16
<i>Validation batch size</i>	16
<i>Weight decay</i>	0.01
<i>Max length</i>	256
<i>Epoch</i>	10

Tabel 4. Konfigurasi Spesifik Hyperparameter

Teknik	<i>Learning rate</i>	r	Lora alpha	Lora dropout
LoRA	$2e-4$	8	16	0.1
<i>Full fine-tuning</i>	$2e-5$	N/A	N/A	N/A

2.5. Metrik Evaluasi

Kinerja semua model dievaluasi menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-Score. Semua metrik ini dihitung menggunakan pendekatan *macro-average*, yang menghitung rata-rata nilai dari masing-masing kelas tanpa mempertimbangkan proporsi jumlah datanya. Selain itu, *confusion matrix* juga digunakan untuk analisis kesalahan yang lebih

terperinci. Secara umum, confusion matrix memiliki empat komponen utama, yaitu *True Positive* (TP) yang merupakan data positif yang berhasil diprediksi sebagai positif, *True Negative* (TN) yang merupakan data negatif yang berhasil diprediksi sebagai negatif, *False Positive* (FP) yang merupakan data negatif yang salah diprediksi sebagai positif, serta *False Negative* (FN) yang merupakan data positif yang salah diprediksi sebagai negatif.

Akurasi = $\frac{TP+TN}{TP+TN+FP+FN}$ Metrik evaluasi dihitung berdasarkan rumus-rumus berikut:

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

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

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

$$\text{F1 - Score} = \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

3. Hasil dan Pembahasan

Hasil dan pembahasan evaluasi model disajikan berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-Score, serta analisis efisiensi penggunaan sumber daya komputasi (VRAM dan durasi pelatihan). Perbandingan komprehensif antara pendekatan *full fine-tuning* dan LoRA juga akan disajikan, diikuti dengan perbandingan performa model dengan hasil dari penelitian sebelumnya. Tabel 5 menunjukkan performa pengenalan emosi dari ketiga model transformer (indobert-large-p2, indobert-base-uncased, dan indobertweet-base-uncased) menggunakan teknik *full fine-tuning* dan LoRA, serta perbandingan dengan penelitian sebelumnya. Sementara itu, Tabel 6 menyajikan perbandingan penggunaan VRAM dan durasi pelatihan antara teknik *full fine-tuning* dan LoRA untuk setiap model transformer yang diuji.

Tabel 5. Perbandingan Performa dengan Penelitian Sebelumnya

Model	Teknik	Aku (%)	Pres (%)	Rec (%)	F1 (%)
indobert-large-p2 (Studi ini)	<i>Full fine-tuning</i>	71,20	69,60	67,12	67,74
	LoRA	71,48	71,09	67,11	67,81
indobert-base-uncased (Studi ini)	<i>Full fine-tuning</i>	70,00	67,02	67,23	67,04
	LoRA	69,72	67,38	66,32	66,51
indobertweet-base-uncased	<i>Full fine-tuning</i>	72,31	70,12	69,51	69,72

(Studi ini)	LoRA	69,44	69,71	64,11	65,42
indobert-large-p2 [22]	<i>Full fine-tuning</i>	65,93	62,24	61,48	65,49
indobert-base-uncased [22]	<i>Full fine-tuning</i>	62,96	60,49	55,57	56,06
indobertweet-base-uncased [22]	<i>Full fine-tuning</i>	64,63	62,62	57,63	58,47
indobertweet-base-uncased + CNN [22]	<i>Full fine-tuning</i>	68,89	65,72	65,43	65,49
indobertweet-base-uncased + BiLSTM + Hyperparameter tuning [22]	<i>Full fine-tuning</i>	66,76	63,13	62,19	62,53
indobertweet-base-uncased + CNN + Hyperparameter tuning [22]	<i>Full fine-tuning</i>	69,26	65,96	65,15	65,49
indobertweet-base-uncased + CNN + <i>weighted approach</i> tuning [22]	<i>Full fine-tuning</i>	69,17	66,19	66,97	66,13

Tabel 6. Perbandingan Penggunaan VRAM dan Durasi Pelatihan

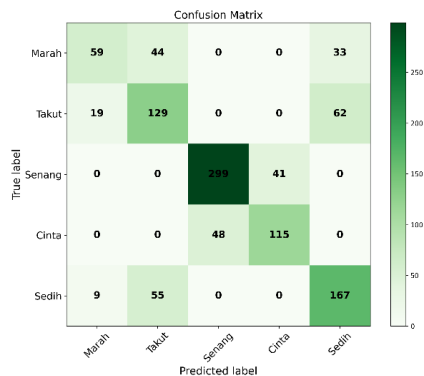
Model	Teknik	VRAM (GB)	Durasi Pelatihan
indobert-large-p2 (Studi ini)	<i>Full fine-tuning</i>	8,8	30:55
	LoRA	7,3	26:54
indobert-base-uncased (Studi ini)	<i>Full fine-tuning</i>	3,3	10:17
	LoRA	2,8	10:33
indobertweet-base-uncased (Studi ini)	<i>Full fine-tuning</i>	3,3	10:03
	LoRA	2,8	09:43

3.1. Analisis Efektivitas *Full fine-tuning* vs. LoRA

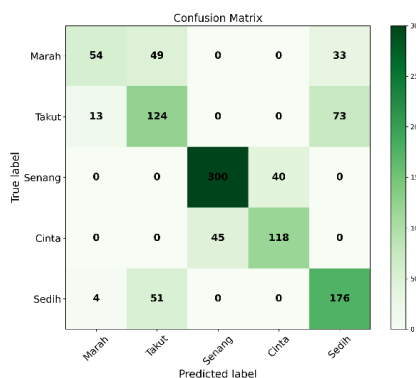
Berdasarkan tabel 5, menunjukkan bahwa pada model indobert-large-p2, teknik LoRA berhasil sedikit mengungguli *full fine-tuning* dalam hal akurasi (71,48% vs 71,20%), presisi (71,09% vs 69,60%), dan F1-score (67,81% vs 67,74%), serta nilai recall pada keduanya juga hampir identik (67,11% vs 67,12%). Hal ini menunjukkan bahwa LoRA tidak hanya kompetitif, tetapi bahkan dapat memberikan sedikit peningkatan kinerja pada model, meskipun hanya melatih sebagian kecil parameter yaitu hanya sekitar 1,05% dari total parameter.

Analisis *confusion matrix* untuk Indobert-large-p2 seperti pada Gambar 2 untuk *full fine-tuning* dan Gambar 3 untuk LoRA, menunjukkan pola prediksi yang serupa. Kedua teknik berhasil dengan baik dalam

mengidentifikasi kelas Senang (*full fine-tuning*: 299 TP, LoRA: 300 TP). Namun, pada kelas Marah, LoRA memiliki *True Positive* (TP) yang sedikit lebih rendah (54 vs 59), tetapi juga *False Positive* (FP) ke “Sedih” yang lebih rendah (4 vs 9), mengindikasikan sensitivitas yang berbeda terhadap kesalahan klasifikasi. Kelas Takut dan Sedih masih menjadi tantangan karena sering kali terjadi kesalahan klasifikasi antar kedua kelas tersebut, meskipun LoRA menunjukkan peningkatan TP untuk “Sedih” (176 vs 167) yang signifikan.

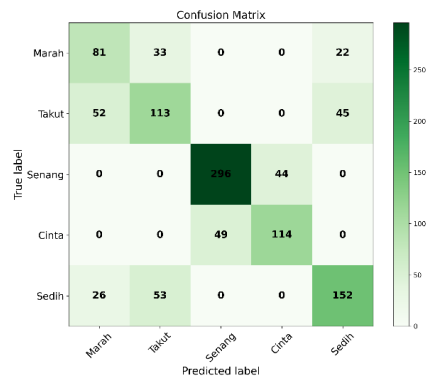


Gambar 2. *Confusion Matrix* Indobert-large-p2 dengan Full Fine-tuning

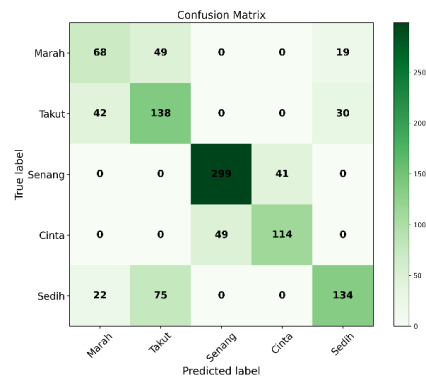


Gambar 3. *Confusion Matrix* Indobert-large-p2 dengan LoRA

Untuk model indobert-base-uncased, *full fine-tuning* sedikit lebih unggul dalam akurasi (70% vs 69,72%) dan F1-score (67,04% vs 66,51%) dibandingkan LoRA, meskipun perbedaan ini tidak terlalu signifikan. Presisi LoRA sedikit lebih tinggi (67,38% vs 67,02%), sementara recall LoRA sedikit lebih rendah. Performa yang relatif setara ini menegaskan kapabilitas LoRA dalam menjaga kualitas prediksi walaupun hanya menggunakan sekitar 1,2% parameter yang dilatih. *Confusion matrix* indobert-base-uncased seperti pada Gambar 4 untuk *full fine-tuning* dan Gambar 5 untuk LoRA, menunjukkan bahwa kelas Senang dan Cinta relatif teridentifikasi dengan baik oleh kedua teknik. Namun, misklasifikasi masih sering terjadi antara Marah, Takut, dan Sedih. Sebagai contoh, kelas Marah sering kali salah diprediksi sebagai Takut atau Sedih.



Gambar 4. *Confusion Matrix* Indobert-base-uncased dengan Full Fine-tuning

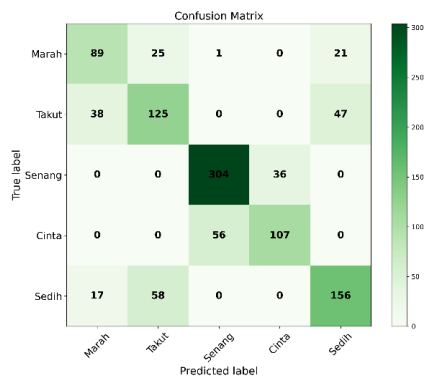


Gambar 5. *Confusion Matrix* Indobert-base-uncased dengan LoRA

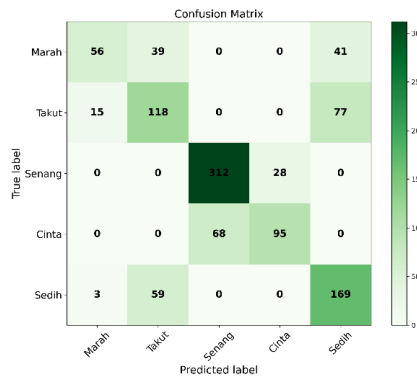
Pada model indobert-base-uncased, *full fine-tuning* kembali menunjukkan performa yang lebih baik secara konsisten di semua metrik dibandingkan LoRA. Akurasi *full fine-tuning* adalah 72,31% berbanding 69,44% untuk LoRA, dengan F1-Score masing-masing 69,72% dan 65,42%. Perbedaan performa tersebut kemungkinan dipengaruhi oleh karakteristik korpus *pre-training* masing-masing model. IndoBERT-large dilatih menggunakan korpus berbahasa Indonesia yang lebih beragam sehingga berpotensi menghasilkan representasi semantik yang lebih sesuai untuk domain ulasan produk. Sebaliknya, IndoBERTweet dikembangkan menggunakan korpus Twitter yang didominasi bahasa informal, singkatan, dan ekspresi khas media sosial. Perbedaan karakteristik korpus tersebut diduga menyebabkan representasi yang dipelajari IndoBERTweet kurang sesuai dengan karakteristik dataset PRDECT-ID yang berasal dari ulasan produk *e-commerce*, sehingga adaptasi menggunakan LoRA belum mampu memberikan performa yang setara dengan *full fine-tuning*. *Confusion matrix* Indobert-base-uncased pada Gambar 4 untuk *full fine-tuning* sedangkan Gambar 5 untuk LoRA, secara lebih lanjut mengilustrasikan perbedaan ini. Meskipun kedua teknik memiliki kekuatan dalam mengidentifikasi “Senang”, LoRA menunjukkan penurunan pada kelas Cinta (hanya 95 TP vs 107 TP pada *full fine-tuning*) dan lebih banyak melakukan misklasifikasi pada kelas Takut dan Sedih, terutama sering salah memprediksi Takut sebagai Sedih, dan sebaliknya. Kelas Marah juga menunjukkan

penurunan TP (56 vs 89 pada *full fine-tuning*), mengindikasikan model LoRA cenderung kurang efektif dalam mengenali emosi tersebut pada hasil model *fine-tuning* Indobertweet ini, walaupun hanya melatih sebagian kecil parameter (1,2% dari total parameter).

Kesalahan klasifikasi yang paling dominan terjadi antara kelas Takut dan Sedih. Hal ini kemungkinan disebabkan oleh kemiripan ekspresi emosional pada ulasan pelanggan, di mana ungkapan seperti kecewa, tidak sesuai, atau menyesal dapat merepresentasikan rasa takut maupun kesedihan sehingga sulit dibedakan oleh model. Selain itu, misklasifikasi antara kelas Marah dan Sedih menunjukkan bahwa beberapa ulasan bernada negatif memiliki karakteristik linguistik yang saling tumpang tindih, sehingga model mengalami kesulitan dalam membedakan kedua kategori emosi tersebut.



Gambar 6. Confusion Matrix indobertweet-base-uncased dengan Full Fine-tuning

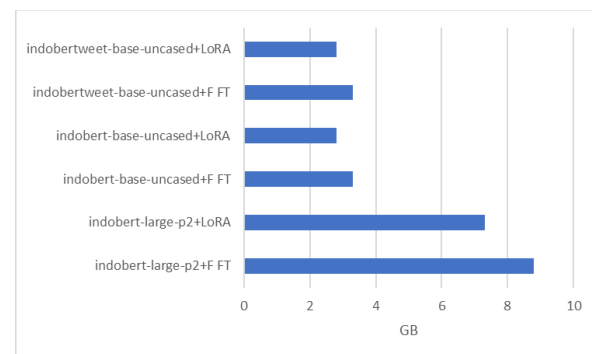


Gambar 7. Confusion Matrix indobertweet-base-uncased dengan LoRA

Secara keseluruhan, meskipun performa LoRA tidak selalu melampaui *full fine-tuning*, perbedaan hasilnya cenderung kecil, dan dalam beberapa kasus, LoRA justru menunjukkan kinerja lebih baik terutama pada model berukuran besar. Temuan ini menunjukkan bahwa teknik LoRA dapat menjadi salah satu metode *fine-tuning* yang sangat kompetitif untuk tugas klasifikasi emosi ulasan produk berbahasa Indonesia.

3.2. Analisis Penggunaan VRAM dan Waktu Pelatihan

Berdasarkan Tabel 6 dan Gambar 8, dapat terlihat pola yang konsisten dalam penggunaan VRAM dan durasi pelatihan antara *full fine-tuning* (F FT) dengan LoRA. LoRA secara signifikan mengurangi kebutuhan VRAM pada semua model. Misalnya, pada model indobert-large-p2, *full fine-tuning* memerlukan 8,8 GB VRAM sedangkan LoRA hanya perlu 7,3 GB. Dengan demikian, LoRA dapat menghemat sebesar 1,5 GB atau sekitar 17%. Untuk model yang lebih kecil seperti indobert-base-uncased dan indobertweet-base-uncased, LoRA juga menunjukkan efisiensi VRAM yang juga dapat mengurangi kebutuhan dari 3,3 GB menjadi 2,8 GB, sehingga terjadi penghematan sekitar 15%. Pengurangan ini sangat krusial, terutama bagi perusahaan kecil dengan dukungan sumber daya komputasi yang terbatas, karena memungkinkan mereka untuk melatih model transformer berukuran besar pada GPU dengan VRAM yang lebih rendah.

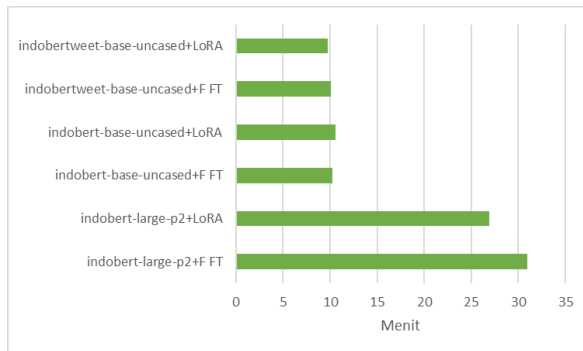


Gambar 8. Perbandingan Penggunaan VRAM antara Full Fine-Tuning dan LoRA

Sementara itu, berdasarkan Tabel 6 dan Gambar 9, selain efisiensi VRAM, LoRA juga berdampak positif pada durasi pelatihan. Pada indobert-large-p2, LoRA berhasil menyelesaikan pelatihan dalam 26 menit 54 detik, lebih cepat sekitar 4 menit dibandingkan *full fine-tuning* yang memerlukan waktu 30 menit 55 detik. Begitu pula dengan indobertweet-base-uncased, LoRA menyelesaikan pelatihan dalam 9 menit 43 detik, lebih cepat sekitar 20 detik dari *full fine-tuning*. Meskipun, pada indobert-base-uncased durasi pelatihan dengan LoRA sedikit lebih lama (10 menit 33 detik) dibandingkan dengan *full fine-tuning* (10 menit 17 detik), perbedaan tersebut sangat kecil sehingga dapat diabaikan.

Efisiensi waktu pelatihan ini penting karena mempercepat siklus iterasi eksperimen dan pengembangan model. Dengan demikian, analisis ini dapat menunjukkan bahwa LoRA dapat menjadi solusi yang sangat efektif dalam mengoptimalkan penggunaan sumber daya komputasi, baik dari segi memori VRAM maupun waktu pelatihan, menjadikannya pilihan yang ideal untuk skenario sumber daya komputasi terbatas. Kemampuan untuk melatih model *state-of-the-art* dengan kebutuhan sumber daya yang lebih rendah dapat menjadi

keuntungan kompetitif yang signifikan bagi perusahaan kecil dalam mengelola analisis emosi dari ulasan pelanggan.



Gambar 9. Perbandingan Durasi Pelatihan antara *Full Fine-Tuning* dan LoRA

3.3. Perbandingan dengan Penelitian Sebelumnya

Perbandingan hasil penelitian ini dengan studi sebelumnya, khususnya oleh Chowanda et al. [22], memberikan gambaran yang jelas mengenai efektivitas pendekatan *fine-tuning* yang digunakan. Berdasarkan Tabel 5, terlihat bahwa studi ini menggunakan model-model Indobert yang sama seperti pada penelitian terdahulu, namun tanpa menambahkan arsitektur kompleks seperti CNN ataupun BiLSTM. Meskipun demikian, model hasil *fine-tuning* dalam penelitian ini justru menunjukkan performa yang secara konsisten lebih unggul. Misalnya, pada model indobert-large-p2, F1-Score yang dicapai melalui *full fine-tuning* sebesar 67,74% dan melalui LoRA sebesar 67,81%, keduanya jauh melampaui hasil dari Chowanda et al. [22], yang hanya mencapai 65,49%. Hal serupa juga ditemukan pada model indobert-base-uncased, di mana studi ini mencatat F1-Score sebesar 67,04% dengan *full fine-tuning* dan 66,51% dengan LoRA, mendapatkan hasil jauh di atas 56,06% yang diperoleh dalam studi sebelumnya. Bahkan pada model indobertweet-base-uncased, hasil penelitian ini menunjukkan F1-Score 69,72% untuk *full fine-tuning* dan 65,42% untuk LoRA, yang secara signifikan lebih tinggi daripada 58,47% dari penelitian Chowanda et al. [22].

Menariknya, meskipun Chowanda et al. [22] mencoba meningkatkan performa dengan menambahkan arsitektur seperti CNN dan BiLSTM, serta menerapkan teknik *hyperparameter tuning* dan pendekatan berbobot (*weighted approach*), hasilnya tetap belum melampaui performa model dasar dalam penelitian ini. Sebagai contoh, model indoberttweet-base-uncased yang dikombinasikan dengan CNN dan *weighted approach tuning* hanya menghasilkan F1-Score sebesar 66,13%, sedangkan model model BiLSTM yang dilengkapi *hyperparameter tuning* bahkan hanya mencapai 62,53%. Angka-angka ini masih berada di bawah performa model-model dalam studi ini, baik yang menggunakan *full fine-tuning* maupun LoRA. Fakta ini menunjukkan bahwa efektivitas *fine-tuning* yang dilakukan terhadap model transformer *pre-*

trained, meskipun hanya dengan teknik *parameter-efficient* seperti LoRA, mampu memberikan hasil yang lebih optimal dibandingkan dengan pendekatan yang memperumit arsitektur model. Keunggulan ini tidak terlepas dari kemampuan model transformer dalam menangkap representasi kontekstual yang kaya, yang bila diadaptasi secara tepat terhadap tugas klasifikasi emosi, terbukti lebih optimal daripada sekadar menambahkan lapisan-lapisan arsitektur tambahan yang belum tentu cocok dengan karakteristik data atau tugasnya. Dengan demikian, penelitian ini menegaskan pentingnya penerapan strategi *transfer learning* yang efisien dan menyoroti potensi besar adaptasi model transformer untuk tugas pengenalan emosi berbahasa Indonesia.

3.4. Batasan dan Penelitian Mendatang

Penelitian ini masih memiliki sejumlah keterbatasan yang dapat menjadi pijakan bagi pengembangan studi di masa mendatang. Salah satu keterbatasan utama adalah bahwa meskipun teknik *Low-Rank Adaptation* (LoRA) menunjukkan efisiensi dalam penggunaan sumber daya, performanya belum secara konsisten melampaui atau menyamai hasil dari *full fine-tuning* pada semua model dan konfigurasi. Hal ini mengindikasikan bahwa penerapan LoRA masih sangat bergantung pada pengaturan hyperparameter yang tepat serta karakteristik model dan data yang digunakan. Oleh karena itu, arah penelitian berikutnya disarankan untuk berfokus pada eksplorasi lebih lanjut terhadap hyperparameter penting dalam LoRA, seperti nilai rank (r), LoRA alpha, dan dropout. Eksperimen yang lebih sistematis terhadap kombinasi parameter ini dapat membantu menemukan konfigurasi optimal yang mampu memaksimalkan kinerja klasifikasi ulasan produk berbahasa Indonesia, sekaligus menjaga efisiensi dalam penggunaan VRAM dan waktu pelatihan.

Selain aspek kuantitatif, penelitian lanjutan juga dapat mencakup analisis kualitatif terhadap kesalahan model, misalnya dengan mengkaji secara manual kasus *false positive* dan *false negative*. Langkah ini penting untuk mengidentifikasi pola kesalahan spesifik yang umum terjadi pada data ulasan produk, seperti kesulitan dalam memahami ironi, metafora, atau ungkapan ambigu yang sekilas tampak netral namun memiliki makna emosional tertentu.

Lebih jauh lagi, eksplorasi terhadap pendekatan *parameter-efficient fine-tuning* lainnya seperti DoRA [36], QLoRA [29], ReFT [37], atau bahkan kombinasi dari beberapa teknik tersebut juga layak untuk dipertimbangkan. Pendekatan ini dapat membantu menjawab pertanyaan apakah terdapat alternatif lain yang mampu memberikan hasil setara atau lebih baik dari LoRA, dengan efisiensi sumber daya yang sama atau bahkan lebih optimal. Dengan demikian, penelitian di masa depan berpeluang besar untuk mengembangkan model yang tidak hanya lebih hemat sumber daya, tetapi juga lebih akurat dan tangguh

dalam menangani kompleksitas ulasan produk berbahasa Indonesia.

4. Kesimpulan

Penelitian ini menyimpulkan bahwa teknik *Low-Rank Adaptation* (LoRA) merupakan pendekatan *fine-tuning* yang kompetitif dan efisien untuk tugas klasifikasi emosi pada ulasan produk berbahasa Indonesia. LoRA mampu mempertahankan performa prediksi yang sebanding dengan *full fine-tuning*, bahkan dalam beberapa kasus seperti pada model indobert-large-p2, LoRA menunjukkan peningkatan F1-score dengan hanya melatih sekitar 1% dari total parameter model. Keunggulan ini disertai dengan efisiensi signifikan dalam penggunaan VRAM dan durasi pelatihan. Pada tiga model transformer yang diuji yaitu indobert-large-p2, indobert-base-uncased, dan indobertweet-base-uncased, LoRA menunjukkan performa yang mendekati atau melampaui *full fine-tuning* dalam metrik utama seperti akurasi, presisi, recall, dan F1-score. Meskipun pada beberapa model seperti indoberttweet-base-uncased, performa LoRA sedikit lebih rendah, perbedaan tersebut bersifat minor dan sebanding dengan penghematan sumber daya komputasi yang diperoleh.

Dari sisi efisiensi, LoRA terbukti dapat mengurangi kebutuhan VRAM sebesar 15-17% dan mempercepat durasi pelatihan beberapa menit dibandingkan *full fine-tuning*. Temuan ini menunjukkan bahwa LoRA dapat menjadi solusi praktis bagi perusahaan kecil dengan keterbatasan sumber daya komputasi. Dengan LoRA, model transformer besar tetap dapat dilatih secara efektif tanpa membutuhkan infrastruktur kelas atas, sehingga membuka peluang adopsi teknologi NLP yang lebih merata di berbagai sektor industri. Selain itu, jika dibandingkan dengan penelitian sebelumnya yang melakukan pendekatan arsitektur model kompleks seperti penambahan CNN atau BiLSTM, LoRA justru menunjukkan bahwa adaptasi model *pre-trained* secara efisien dapat menghasilkan performa yang lebih unggul. Oleh karena itu, strategi *parameter-efficient fine-tuning* seperti LoRA tidak hanya menawarkan solusi hemat sumber daya, tetapi juga mampu meningkatkan kapabilitas pengenalan emosi ulasan produk dalam berbahasa Indonesia.

Reference

- [1] Y. M. Ginting, T. Chandra, I. Miran, and Y. Yusriadi, "Repurchase intention of e-commerce customers in Indonesia: An overview of the effect of e-service quality, e-word of mouth, customer trust, and customer satisfaction mediation," *International Journal of Data and Network Science*, vol. 7, no. 1, pp. 329–340, 2023, doi: 10.5267/j.ijdns.2022.10.001.
- [2] S. Nida, A. Nurhakim, J. M. Noor Isiqamah, and Nuraini, "ANALISIS PERKEMBANGAN TOKO ONLINE (E-COMMERCE) DI INDONESIA," *Jurnal Bisnis Digital*, vol. 2, no. 1, pp. 126–137, May 2024, doi: 10.52060/j-bisdig.v2i1.2180.
- [3] Rachmiani, N. K. Oktadinna, and T. R. Fauzan, "The Impact of Online Reviews and Ratings on Consumer Purchasing Decisions on E-commerce Platforms," *International Journal of Management Science and Information Technology*, vol. 4, no. 2, pp. 504–515, Dec. 2024, doi: 10.35870/ijmsit.v4i2.3373.
- [4] Y. Sun, K. Sekiguchi, and Y. Ohsawa, "The Impact of Sentiment Scores Extracted from Product Descriptions on Customer Purchase Intention," *New Gener. Comput.*, vol. 42, no. 4, pp. 617–633, 2024, doi: 10.1007/s00354-024-00242-9.
- [5] A. Ananda and A. Sudrajat, "Pengaruh Ulasan Produk dan Layanan Purna Jual terhadap Minat Beli Ulang Smartphone Xiaomi Redmi Note 11 (Survei pada Komunitas Pengguna Xiaomi Redmi Note 11)," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 15, Sep. 2022, doi: 10.5281/zenodo.7040099.
- [6] S. Sudaryanto, H. Anifatul, R. D. Ivana, K. Asila Dwi, and R. and Rusdiyanto, "The mediating effect of customer trust of E-WOM and online customer reviews impacting purchase decision of household electronic products at a marketplace: evidence from Indonesia," *Cogent Business & Management*, vol. 12, no. 1, p. 2503093, Dec. 2025, doi: 10.1080/23311975.2025.2503093.
- [7] A. B. M. Samiaji, D. A. Pramesti, and M. W. Ibrahim, "Pengaruh Kualitas Produk, Ulasan Konsumen, dan Citra Merek terhadap Kepuasan Pelanggan Melalui Keputusan Pembelian," in *Prosiding Business and Economics Conference in Utilizing of Modern Technology (BIS-HSS 2024)*, in UMMagelang Conference Series. Universitas Muhammadiyah Magelang, Aug. 2024, doi: 10.31603/conference.12041.
- [8] L. Livina and H. K. Tunjungsari, "Pengaruh persepsi kegunaan dari ulasan online, kepercayaan konsumen, dan persepsi risiko pada intensi membeli produk busana secara online," *Jurnal Manajemen Bisnis dan Kewirausahaan*, vol. 9, no. 2, pp. 352–364, Mar. 2025, doi: 10.24912/jmbk.v9i2.32361.
- [9] M. Sun and J. Zhao, "Behavioral Patterns beyond Posting Negative Reviews Online: An Empirical View," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 3, pp. 949–983, 2022, doi: 10.3390/jtaer17030049.
- [10] M. F. A. Thariq, "Pengaruh E-WOM, Customer Relationship Management, dan Emotional Branding Terhadap Loyalty Konsumen," *Jurnal Manajemen Pemasaran Dan Perilaku Konsumen*, vol. 2, no. 3, pp. 814–827, 2023, doi: 10.21776/jmpmk.2022.02.3.23.
- [11] A. Rahmawati, D. Sugandini, and Y. Istanto, "Pengaruh Customer Experience terhadap Attitude Loyalty dan Behavioral Loyalty yang Dimediasi oleh Emotional Experience pada Pengguna Mobile Application Shopee (Studi Kasus di Yogyakarta)," *JMBI UNSRAT (Jurnal Ilmiah Manajemen Bisnis dan Inovasi Universitas Sam Ratulangi)*, vol. 8, no. 3, Dec. 2021, doi: 10.35794/jmbi.v8i3.36740.
- [12] S. Kundu and S. Chakraborti, "A Comparative Study of Online Consumer Reviews Among Mainstream and Niche Products: Analysis Across Emerging and Developed Markets," *International Journal of Business Analytics*, vol. 11, no. 1, 2024, doi: 10.4018/IJBAN.353306.
- [13] Y. K. Dwivedi et al., "Setting the future of digital and social media marketing research: Perspectives and research propositions," *Int. J. Inf. Manage.*, vol. 59, p. 102168, 2021, doi: https://doi.org/10.1016/j.ijinfomgt.2020.102168.
- [14] A. Zablocki, K. Makri, and M. J. Houston, "Emotions Within Online Reviews and their Influence on Product Attitudes in Austria, USA and Thailand," *Journal of Interactive Marketing*, vol. 46, pp. 20–39, 2019, doi: https://doi.org/10.1016/j.intmar.2019.01.001.
- [15] G. Kaur and A. Sharma, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-022-00680-6.
- [16] A. Devi Putri Ariyanto, F. K. Fikriah, and A. F. Setyawan, "Impact of Statistical and Semantic Features Extraction for Emotion Detection on Indonesian Short Text Sentences," *Commit Journal*, vol. 19, no. 1, pp. 1–13, 2025, doi: 10.21512/commit.v19i1.11680.
- [17] M. A. Riza and N. Charibaldi, "Emotion Detection in Twitter Social Media Using Long Short-Term Memory (LSTM) and Fast Text," *International Journal of Artificial Intelligence &*

- Robotics (IJAIR), vol. 3, no. 1, pp. 15–26, 2021, doi: 10.25139/ijair.v3i1.3827.
- [18] T. Chutia and N. Baruah, “A review on emotion detection by using deep learning techniques,” *Artif. Intell. Rev.*, vol. 57, no. 8, p. 203, 2024, doi: 10.1007/s10462-024-10831-1.
- [19] M. S. Saputri, R. Mahendra, and M. Adriani, “Emotion Classification on Indonesian Twitter Dataset,” in 2018 International Conference on Asian Language Processing (IALP), 2018, pp. 90–95. doi: 10.1109/IALP.2018.8629262.
- [20] D. Haryadi and G. P. Kusuma, “Emotion Detection in Text using Nested Long Short-Term Memory,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019, doi: 10.14569/IJACSA.2019.0100645.
- [21] T. Liu, Y. Du, and Q. Zhou, “Text Emotion Recognition Using GRU Neural Network with Attention Mechanism and Emotion Emotions,” in *Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence*, in RICAI '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 278–282. doi: 10.1145/3438872.3439094.
- [22] A. Chowanda, R. Sutoyo, S. Achmad, E. W. Andangsari, S. M. Isa, and T.-K. Chen, “MODELING EMOTIONS RECOGNITION ON INDONESIAN PRODUCT REVIEW BY COMBINING BERT, CNN, AND LSTM ARCHITECTURE,” *International Journal of Innovative Computing Information and Control*, vol. 20, no. 3, pp. 929–944, 2024, doi: 10.24507/ijicic.20.03.929.
- [23] R. Sutoyo, S. Achmad, A. Chowanda, E. W. Andangsari, and S. M. Isa, “PRDECT-ID: Indonesian product reviews dataset for emotions classification tasks,” *Data Brief*, vol. 44, 2022, doi: 10.1016/j.dib.2022.108554.
- [24] Y. O. Sihombing, R. Fuad Rachmadi, S. Sumpeno, and Moh. J. Mubarak, “Optimizing IndoRoBERTa Model for Multi-Class Classification of Sentiment & Emotion on Indonesian Twitter,” in 2024 IEEE 10th Information Technology International Seminar (ITIS), 2024, pp. 12–17. doi: 10.1109/ITIS64716.2024.10845566.
- [25] H. Ahmadian, T. F. Abidin, H. Riza, and K. Muchtar, “Transformer-Based Indonesian Language Model for Emotion Classification and Sentiment Analysis,” in 2023 International Conference on Information Technology and Computing (ICITCOM), 2023, pp. 209–214. doi: 10.1109/ICITCOM60176.2023.10442970.
- [26] L. Wang et al., “Parameter-efficient fine-tuning in large language models: a survey of methodologies,” *Artif. Intell. Rev.*, vol. 58, no. 8, p. 227, 2025, doi: 10.1007/s10462-025-11236-4.
- [27] N. Houlsby et al., “Parameter-Efficient Transfer Learning for NLP,” *CoRR*, vol. abs/1902.00751, 2019, [Online]. Available: <http://arxiv.org/abs/1902.00751>
- [28] X. L. Li and P. Liang, “Prefix-Tuning: Optimizing Continuous Prompts for Generation,” *CoRR*, vol. abs/2101.00190, 2021, [Online]. Available: <https://arxiv.org/abs/2101.00190>
- [29] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, “QLoRA: efficient finetuning of quantized LLMs,” in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, in NIPS '23. Red Hook, NY, USA: Curran Associates Inc., 2023.
- [30] E. J. Hu et al., “LoRA: Low-Rank Adaptation of Large Language Models,” *CoRR*, vol. abs/2106.09685, 2021, [Online]. Available: <https://arxiv.org/abs/2106.09685>
- [31] V. B. Parthasarathy, A. Zafar, A. Khan, and A. Shahid, “The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities,” 2024. [Online]. Available: <https://arxiv.org/abs/2408.13296>
- [32] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” *CoRR*, vol. abs/2011.00677, 2020, [Online]. Available: <https://arxiv.org/abs/2011.00677>
- [33] F. Koto, J. H. Lau, and T. Baldwin, “IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, M.-F. Moens, X. Huang, L. Specia, and S. W. Yih, Eds., Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 10660–10668. doi: 10.18653/v1/2021.emnlp-main.833.
- [34] S. Hayou, N. Ghosh, and B. Yu, “LoRA+: Efficient Low Rank Adaptation of Large Models,” 2024. [Online]. Available: <https://arxiv.org/abs/2402.12354>
- [35] D. Biderman et al., “LoRA Learns Less and Forgets Less,” 2024. [Online]. Available: <https://arxiv.org/abs/2405.09673>
- [36] S.-Y. Liu et al., “DoRA: weight-decomposed low-rank adaptation,” in *Proceedings of the 41st International Conference on Machine Learning*, in ICML'24. JMLR.org, 2024.
- [37] Z. Wu et al., “ReFT: Representation Finetuning for Language Models,” 2024. [Online]. Available: <https://arxiv.org/abs/2404.03592>