

# Opinion Mining of Pedometer Application Reviews on Google Play Store Using Fine-Tuned IndoBERT-Base

Anggi Setyo Primono <sup>1\*</sup>, Ucta Pradema Sanjaya <sup>2\*</sup>

\* Informatics Engineering, Ngudi Waluyo University  
[anggiunw@gmail.com](mailto:anggiunw@gmail.com)<sup>1</sup>, [uctapradema@unw.ac.id](mailto:uctapradema@unw.ac.id)<sup>2</sup>

## Article Info

### Article history:

Received 2026-01-06

Revised 2026-01-23

Accepted 2026-02-11

### Keyword:

*Opinion Mining,  
Sentiment Analysis,  
IndoBERT,  
Fine-Tuning,  
Google Play Store,  
Pedometer Application.*

## ABSTRACT

User reviews on the Google Play Store provide valuable insights into user satisfaction and application performance. However, manual analysis of these reviews is inefficient due to large data volume and the informal characteristics of the Indonesian language. This study proposes an opinion mining approach using a fine-tuned IndoBERT-Base model to classify user sentiments into three classes: positive, neutral, and negative. A total of 1,665 reviews of a Pedometer application were collected, with 1,636 reviews retained after preprocessing. The dataset was divided into training, validation, and test sets using stratified sampling to preserve class distribution. Experimental results show that the proposed model achieves an accuracy of 94.51% and a weighted F1-score of 0.93 on the test set. Despite strong overall performance, the results indicate that class imbalance significantly affects the classification of neutral and negative sentiments. Error analysis reveals that ambiguous expressions and limited samples in minority classes remain challenging for the model. This study demonstrates that fine-tuned IndoBERT-Base is effective for sentiment analysis of Indonesian mobile application reviews while highlighting the importance of addressing imbalanced data in opinion mining tasks.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. PENDAHULUAN

Pesatnya kemajuan teknologi seluler memicu tren pengembangan perangkat lunak di sektor kesehatan, khususnya pada kategori aplikasi pedometer. Kepopuleran aplikasi ini didorong oleh kemudahannya dalam membantu pengguna melacak aktivitas fisik harian, termasuk perhitungan langkah kaki secara *real-time*. Di sisi lain, *Google Playstore* berperan sebagai saluran komunikasi krusial yang menampung beragam opini dan ulasan pengguna. Kumpulan data tekstual tersebut merupakan aset strategis bagi para pengembang untuk mengevaluasi kinerja produk serta memetakan kebutuhan *user* secara lebih akurat.

Volume ulasan pengguna pada *Google Playstore* menyajikan data opini yang sangat masif sekaligus heterogen. Tantangan utamanya terletak pada metodologi pengelolaan dan ekstraksi informasi mendalam—seperti polaritas sentimen dan poin kritik spesifik—dari sekumpulan data teks tidak terstruktur secara efektif. Saat ini, para pengembang aplikasi pedometer masih terkendala oleh keterbatasan sistem

otomatis yang mampu mengidentifikasi sentimen mayoritas secara akurat. Ketiadaan instrumen klasifikasi yang andal untuk memetakan kategori positif, negatif, maupun netral menghambat proses evaluasi fitur aplikasi secara cepat dan tepat.

Studi mengenai analisis sentimen, atau yang secara teknis dikenal sebagai penambangan opini (*opinion mining*), kini menjadi fokus utama dalam ranah rekayasa perangkat lunak (*software engineering*). Para akademisi di bidang ini telah memanfaatkan teknik tersebut untuk berbagai tujuan strategis, mulai dari memetakan kondisi emosional pengembang pada komentar sumber kode hingga mengekstraksi poin-poin krusial dari umpan balik pengguna aplikasi seluler. Namun, pesatnya pertumbuhan literatur di bidang ini menghadirkan tantangan tersendiri; peneliti sering kali membutuhkan waktu yang tidak sedikit untuk menentukan metodologi yang paling relevan serta memahami potensi risiko teknis yang mungkin timbul dari penerapan pendekatan tersebut. [1].

Walaupun kolom ulasan di *Google Playstore* menyajikan data komprehensif terkait keunggulan dan defisiensi aplikasi

pedometer, besarnya volume data serta format teks yang tidak terstruktur menjadi kendala utama dalam analisis manual. Proses penarikan kesimpulan secara konvensional tidak hanya memakan waktu lama, tetapi juga rentan terhadap subjektivitas peneliti. Oleh karena itu, diperlukan sebuah pendekatan otomatisasi yang memiliki tingkat akurasi tinggi dan bersifat objektif guna mengekstraksi polaritas sentimen pengguna. Kebutuhan ini menjadi semakin krusial terutama dalam pemrosesan ulasan berbahasa Indonesia yang memiliki karakteristik linguistik yang khas dan beragam.

Penambahan opini merupakan salah satu implementasi *Natural Language Processing* (NLP) yang paling krusial dalam mengekstraksi intensi atau maksud manusia melalui ulasan dan tanggapan tertulis. Di ranah pendidikan, teknologi ini dimanfaatkan secara strategis untuk menangkap aspirasi siswa guna menyempurnakan praktik pedagogis secara berkelanjutan. Berkat inovasi pada teknik anotasi serta metodologi Kecerdasan Buatan (AI), proses pelabelan sentimen (positif, negatif, dan netral) pada komentar siswa kini dapat dilakukan secara otomatis dengan intervensi manual yang minimal. Artikel ini bertujuan untuk mengeksplorasi kontribusi analisis emosional dalam dunia pendidikan yang ditinjau melalui empat tingkatan utama: dokumen, kalimat, entitas, serta aspek. [2].

Secara khusus, artikel ini menyajikan tinjauan komprehensif mengenai *Opinion Mining* (PO), yang mencakup pemetaan teknologi terkini serta berbagai tantangan implementatif guna memberikan gambaran ruang lingkup penelitian yang jelas bagi para akademisi. Penulisan ini dirancang sebagai panduan fundamental bagi peneliti pemula untuk memperoleh pemahaman utuh mengenai alur kerja PO. Lebih jauh lagi, hasil tinjauan ini memiliki relevansi tinggi bagi divisi riset di berbagai organisasi yang berfokus pada pengolahan umpan balik pelanggan. Intisari dari studi ini adalah mengompilasi serta menyajikan teknik-teknik PO yang paling dominan serta tren inovasi terbaru dalam disiplin ilmu tersebut. [3].

Pedometer merupakan instrumen portabel, baik dalam bentuk elektronik maupun elektromekanis, yang dikembangkan secara spesifik untuk melakukan kalkulasi terhadap jumlah langkah kaki individu. Secara teknis, perangkat ini beroperasi dengan mendeteksi dinamika pergerakan tubuh, khususnya melalui sensor yang diposisikan pada area pinggul atau pergelangan tangan. Selain berfungsi sebagai alat pemantau intensitas aktivitas fisik harian, pedometer juga berperan dalam penetapan target kebugaran—seperti standar 10.000 langkah—guna menstimulasi gaya hidup sehat. Melalui pengolahan data langkah tersebut, perangkat ini juga mampu mengestimasi variabel tambahan seperti jarak tempuh serta total pembakaran kalori pengguna.

Kelebihan utama dari program intervensi kesehatan yang memanfaatkan teknologi pedometer terletak pada signifikansi otonomi partisipan, di mana individu tidak lagi terkekang oleh protokol jadwal latihan yang rigid maupun ketergantungan pada fasilitas olahraga tertentu. Pendekatan ini secara

strategis dirancang untuk menguatkan motivasi intrinsik subjek melalui sistem insentif mandiri yang berorientasi pada pencapaian ambang batas langkah harian yang telah dikalibrasi. Kebebasan penuh yang diberikan dalam menentukan lokus serta durasi kegiatan menjadikan aktivitas berjalan kaki sebagai pilihan yang paling realistis dan dominan bagi kelompok populasi dewasa dengan tingkat mobilitas rendah. Melalui integrasi sistem pelacakan langkah yang berkelanjutan, intervensi ini tidak hanya memantau progres fisik secara akurat, tetapi juga mendorong transformasi gaya hidup yang lebih aktif dan mandiri tanpa intervensi eksternal yang bersifat restriktif. [4].

Baharuddin dan Naufal (2023) meneliti optimalisasi model *IndoBERT pre-trained* melalui tahap *fine-tuning* untuk klasifikasi soal Bahasa Indonesia berdasarkan taksonomi *Bloom*. Langkah ini diambil guna mengatasi kendala kategorisasi manual yang tidak efisien. Melalui penyesuaian *hyperparameter*, studi tersebut berupaya meningkatkan standar presisi dan akurasi model, yang kemudian dievaluasi menggunakan metrik performa komprehensif seperti akurasi, *F1-score*, *precision*, dan *recall*. [5].

Penelitian ini masih menyisakan keterbatasan, terutama pada kesulitan model *IndoBERT* dalam mendiferensiasi sentimen netral dan negatif pada teks multi-aspek. Selain itu, intensitas emosi serta hubungan antar-aspek dalam *User Generated Content* (UGC) belum terpetakan secara optimal. Hal ini mengindikasikan perlunya pengembangan strategi segmentasi aspek dan pengayaan arsitektur model guna meningkatkan akurasi klasifikasi pada domain dengan karakteristik bahasa informal yang kompleks. [6].

Di sisi lain, studi yang dilakukan oleh Simanjuntak et al. (2024) memberikan perspektif kritis mengenai implementasi *IndoBERT*, di mana efektivitas model dalam mendeteksi berita bohong (*fake news*) terbukti sangat dipengaruhi oleh konfigurasi *hyperparameter* pada fase *fine-tuning*. Penelitian tersebut menekankan bahwa tanpa optimasi yang presisi, performa model tidak akan mencapai titik maksimal. Guna membuktikan hal tersebut, mereka melakukan eksperimen komparatif menggunakan berbagai metodologi optimasi, meliputi *Bayesian optimization*, *grid search*, hingga *random search*. Hasil evaluasi menunjukkan bahwa pendekatan berbasis *Bayesian optimization* memberikan keunggulan performa paling signifikan, yang tercermin pada tingginya nilai *precision*, *recall*, dan *F1-score*. Temuan ini menyingkap adanya celah penelitian (*research gap*) terkait kebutuhan akan evaluasi sistematis terhadap pengaruh konfigurasi teknis pada arsitektur *IndoBERT*. Lebih jauh lagi, studi ini membuka jalan bagi pengembangan teknik optimasi yang lebih adaptif terhadap keunikan dataset tertentu, sehingga kemampuan generalisasi model dapat diperkuat tidak hanya pada ranah klasifikasi hoaks, tetapi juga pada berbagai tugas *Natural Language Processing* (NLP) lainnya secara lebih luas. [7].

Fokus utama dari penelitian ini adalah mengimplementasikan arsitektur *IndoBERT-Base* sebagai instrumen cerdas untuk mengklasifikasikan polaritas sentimen pada sekumpulan ulasan aplikasi Pedometer yang

bersumber dari Google Play Store. Melalui penerapan strategi *fine-tuning* yang mendalam, studi ini diupayakan mampu menghasilkan sebuah model analisis sentimen dengan tingkat presisi tinggi yang dapat membantu pengembang aplikasi dalam mengekstraksi kebutuhan serta aspirasi pengguna secara otomatis dan efisien. Lebih dari sekadar pencapaian akurasi teknis, luaran dari penelitian ini diproyeksikan menjadi referensi ilmiah yang signifikan bagi para akademisi di masa mendatang mengenai optimalisasi penggunaan model bahasa pra-latih (*pretrained model*) bahasa Indonesia. [8].

Dengan memanfaatkan algoritma, penambangan opini menerapkan penilaian berbasis aturan, otomatis, atau gabungan untuk secara otomatis mengklasifikasikan kata-kata pelanggan sebagai positif, negatif, atau netral. Lebih dari sekadar klasifikasi sentimen, teknik ini mampu mengekstraksi polaritas (intensitas positif / negatif), subjek, dan pemegang opini yang relevan dalam teks. Metode ini fleksibel digunakan untuk menganalisis segmen teks apa pun, mulai dari dokumen lengkap hingga potongan kecil seperti kalimat [9].

Implementasi aplikasi ini didasarkan pada pemanfaatan sensor pedometer terintegrasi pada perangkat seluler guna mencapai akurasi dalam penghitungan langkah. Hal ini memberikan kemudahan bagi pengguna untuk memantau metrik langkah, melacak progres secara berkala (harian, mingguan, bulanan), serta mengoptimalkan pengaturan target aktivitas fisik [10].

Studi sebelumnya telah menerapkan pendekatan pembelajaran mesin dan pembelajaran mendalam seperti SVM, LSTM, dan IndoBERT untuk analisis sentimen teks bahasa Indonesia. Dibandingkan dengan pendekatan-pendekatan tersebut, IndoBERT menunjukkan pemahaman kontekstual yang lebih unggul, khususnya untuk bahasa informal yang umum ditemukan dalam ulasan aplikasi.

Studi ini memberikan kontribusi pada literatur yang ada dalam tiga aspek utama. Pertama, studi ini memberikan evaluasi empiris dari IndoBERT-Base yang telah disempurnakan untuk klasifikasi sentimen multikelas pada ulasan aplikasi seluler Indonesia, khususnya di bidang kesehatan dan kebugaran. Kedua, penelitian ini menyoroti dampak ketidakseimbangan kelas yang parah pada kinerja model dengan menganalisis metrik per kelas dan matriks kebingungan, yang sering diabaikan dalam studi sebelumnya yang terutama melaporkan akurasi. Ketiga, analisis kesalahan dilakukan untuk mengidentifikasi pola linguistik yang menyebabkan kesalahan klasifikasi, terutama antara sentimen netral dan negatif. Kontribusi ini membedakan studi ini dari penelitian analisis sentimen sebelumnya yang terutama berfokus pada metrik kinerja keseluruhan tanpa interpretasi kesalahan yang mendalam.

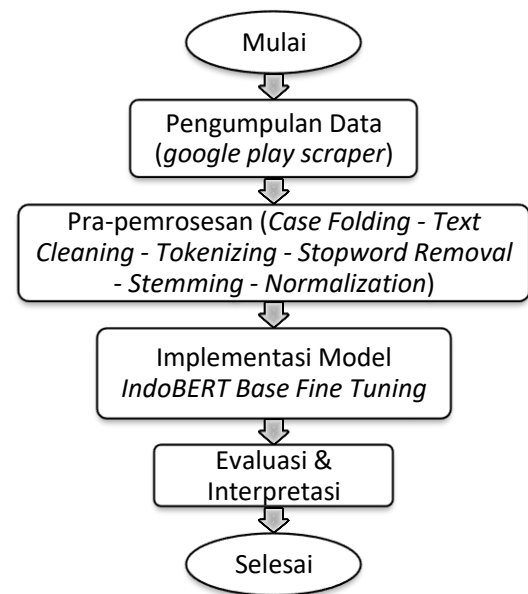
## II. METODE

Bab ini menguraikan secara sistematis pendekatan dan langkah-langkah yang digunakan dalam penelitian *Opinion Mining of Pedometer Application Reviews on Google Play Store Using Fine Tuned IndoBERT Base*.

Penelitian ini menggunakan pendekatan kuantitatif untuk mengekstrak, memproses, dan menganalisis ulasan pengguna dari aplikasi pedometer yang tersedia di *Google Playstore*.

Kerangka kerja penelitian ini dirancang untuk mencapai tujuan mengklasifikasikan sentimen ulasan (positif, negatif, atau netral) dan mendapatkan wawasan menyeluruh mengenai opini pengguna terhadap berbagai aspek aplikasi.

Metode yang dipilih, yaitu *IndoBERT Base Fine Tuning*, merupakan modifikasi dari analisis sentimen berbasis IndoBERT standar yang diperkuat dengan prinsip statistik *Fine Tuning* untuk penentuan skor polaritas, menjadikannya sesuai untuk menangani kompleksitas bahasa alami yang ditemukan dalam ulasan pengguna.



Gambar 1. Diagram alir penelitian

Sub-bab selanjutnya akan merinci sumber data, tahapan pra-pemrosesan data (*pre-processing*), implementasi model analisis sentimen, hingga metrik evaluasi yang digunakan.

### A. Tahap Pengumpulan Data

Pengumpulan data merupakan komponen inti dan prosedur mendasar dalam kerangka metodologi penelitian yang sistematis. Proses ini secara inheren melibatkan objek, subjek, instrumen, dan teknik yang spesifik. Teknik pengumpulan data sendiri adalah pendekatan strategis yang dipilih peneliti untuk mendapatkan informasi atau data penelitian yang relevan.

Agar implementasi teknik dan instrumen penelitian menjadi efektif, peneliti harus menjalankan serangkaian langkah operasional, meliputi: identifikasi informan dan lokasi studi, pengecekan kemudahan akses terhadap sumber data, penentuan jenis informasi yang diperlukan untuk akuisisi data, perancangan protokol dan perangkat penelitian yang instrumental, pengadministrasian data penelitian yang telah berhasil dihimpun.

Keabsahan data (*data validity*) adalah faktor penentu yang menjamin akurasi cerminan fenomena yang diteliti, meminimalkan potensi bias, memfasilitasi generalisasi hasil, serta mendukung pengambilan keputusan yang tepat. Oleh karena itu, analisis keabsahan sangat vital untuk mempertahankan kredibilitas, reliabilitas, dan validitas hasil penelitian ilmiah [11].

Proses pengumpulan data dilakukan menggunakan pendekatan *web scraping*. Perangkat lunak yang diimplementasikan adalah paket *Google-play-scrapers* yang dijalankan melalui lingkungan pemrograman Python. Ulasan pengguna aplikasi Pedometer yang berhasil diekstraksi dari *Google Play Store* memiliki format kalimat atau paragraf yang selanjutnya akan dianalisis [12].

**B. Tahap Pra-pemrosesan**

Pra-pemrosesan data merupakan tahapan krusial dalam implementasi algoritma pembelajaran mesin (*machine learning*) dan analisis teks. Fungsi utama dari tahap ini adalah untuk membersihkan, menormalisasi, dan mempersiapkan data mentah agar siap dianalisis, yang pada akhirnya akan meningkatkan akurasi dan kualitas hasil yang diperoleh.

Secara spesifik, pra-pemrosesan data mencakup serangkaian operasi, termasuk *case folding* (penyeragaman huruf), *text cleaning* (pembersihan teks), *tokenizing* (pembagian kata), *Stopword Removal*, *stemming / Lemmatization* (normalisasi kata), *normalization* (normalisasi khusus) dan pelabelan.



Gambar 2. Tahap Pra-pemrosesan

Tahapan ini sangat esensial sebagai prasyarat dalam penerapan berbagai algoritma pembelajaran mesin, termasuk model IndoBERT yang digunakan untuk klasifikasi [13].

1) *Case Folding*. Tahap ini dilakukan untuk memastikan keseragaman representasi huruf pada data teks. Proses ini merupakan tahap awal pra-pemrosesan yang esensial guna menghindari ambiguitas dan perbedaan interpretasi antar kata akibat variasi kapitalisasi [14].

2) *Text Cleanin*. Langkah ini dilaksanakan sebagai upaya preprocessing untuk memurnikan data dari artefak-artefak yang tidak relevan dan anomali. Selain itu, proses ini juga mencakup mekanisme untuk mendegradasi atau menghapus entri yang terduplikasi (*duplicate entries*) sehingga menjamin integritas dan keunikan dataset [15].

3) *Tokenizing* adalah proses mengubah corpus teks menjadi segmen-segmen kecil yang lazimnya berupa kata (*token*). Melalui langkah ini, peneliti dapat mengidentifikasi dan mengekstrak setiap komponen leksikal dalam dataset, menjadikannya dasar untuk pemrosesan teks selanjutnya [16].

4) *Stopword Removal*, merupakan tahapan pra-pemrosesan teks yang esensial dalam analisis teks dan pemrosesan bahasa alami. Prosedur ini bertujuan untuk mengeliminasi kata-kata dengan frekuensi kemunculan tinggi (*common words*) yang tidak memberikan kontribusi makna yang substansial. Contoh *stopword* dalam suatu bahasa (misalnya, "dan," "di," "yang" dalam Bahasa Indonesia) dianggap tidak signifikan untuk ekstraksi informasi atau proses penambangan teks [17].

5) *Stemming*, diartikan sebagai prosedur dalam text preprocessing yang berfungsi untuk mereduksi variasi morfologis dari suatu kata ke bentuk dasar atau akarnya (*root form*). Proses ini dilaksanakan dengan cara mengeliminasi semua afiks (imbuan—prefiks, sufiks, infiks, atau konfiks) yang melekat pada kata turunan, sehingga menghasilkan kata dasar yang merepresentasikan makna leksikal utuh [18].

6) *Normalization*, merupakan teknik transformasi linear yang diaplikasikan untuk mitigasi permasalahan hubness dalam konteks Pembelajaran Few-Shot (FSL). Prosedur ini melibatkan standarisasi di mana setiap komponen pada vektor fitur dikurangi dengan nilai rata-rata (*mean*) lalu dibagi dengan deviasi standar (*standard deviation*) dari seluruh komponen vector [19].

**C. Tahap Implementasi Model**

Implementasi model adalah tahap krusial dalam sebuah penelitian atau proyek yang berfokus pada menerapkan model atau algoritma yang telah dirancang ke data aktual untuk menghasilkan luaran (*output*).

1) *Instalasi Pustaka Hugging Face* : Hugging Face sebenarnya adalah sebuah ekosistem, jadi "instalasinya" tergantung pada pustaka (*library*) mana yang ingin digunakan. Namun, yang paling umum dan utama adalah pustaka *transformers*. Langkah awal persiapan lingkungan ini Sangat disarankan untuk menggunakan Virtual Environment agar tidak bentrok dengan pustaka lain di komputer. [20].

2) *Penghitungan Probabilitas Sentimen Dokumen (Opinion)*: Metode ini menggunakan IndoBERT Base untuk menghitung probabilitas sebuah ulasan D diklasifikasikan ke dalam kelas sentimen C (Positif, Negatif, atau Netral). Rumus umum:

$$P(C|D) = \frac{P(D|C).P(C)}{P(D)} \dots\dots (1)$$

Penerapan pada teks dengan mengingat ulasan D terdiri dari serangkaian kata  $w_1, w_2, \dots, w_n$ , probabilitas  $P(D|C)$  dihitung sebagai produk dari probabilitas kemunculan setiap kata, dengan asumsi (kata-kata saling independen):

$$P(C|D) \approx P(w_1|C) \cdot P(w_2|C) \dots P(w_n|C) \quad \dots\dots (2)$$

3) *Klasifikasi Sentimen* : Langkah Klasifikasi Sentimen ulasan D ditentukan oleh kelas C yang memberikan probabilitas tertinggi ( $\text{argmax}_C P(C|D)$ ). Setiap ulasan akan melalui proses perhitungan probabilitas, dan hasilnya akan ditetapkan sebagai salah satu kelas sentimen:

Positif: Jika  $P(\text{Positif} | D)$  adalah yang tertinggi.

Negatif: Jika  $P(\text{Negatif} | D)$  adalah yang tertinggi.

Netral: Jika selisih probabilitas antar kelas sangat kecil, atau sesuai dengan ambang batas yang ditetapkan.

4) *Agregasi dan Visualisasi Hasil* : Agregasi dengan hitung jumlah total ulasan untuk setiap kelas sentimen (Positif, Negatif, Netral) dari keseluruhan data atau per aplikasi pedometer. Visualisasi dengan menyajikan hasil dalam bentuk yang mudah dipahami (misalnya, diagram lingkaran atau batang) untuk menunjukkan distribusi sentimen secara keseluruhan.

#### D. Tahap Evaluasi

Tahap Evaluasi adalah langkah krusial untuk mengukur kinerja dan keandalan dari model Bayesian Lexicon yang telah diimplementasikan dalam klasifikasi sentimen ulasan aplikasi pedometer. Tujuannya adalah memastikan bahwa hasil *opinion mining* yang diperoleh adalah akurat, valid, dan dapat dipertanggungjawabkan secara ilmiah.

1) *Metrik Evaluasi Kinerja* : Untuk menilai seberapa baik model mengklasifikasikan sentimen (Positif, Negatif, Netral), menggunakan metrik evaluasi standar. Ini dilakukan dengan membandingkan hasil klasifikasi model (*Prediction*) dengan label yang benar (*True Label*) dari sebagian kecil data yang sudah dilabeli (*Data Ground Truth*). Metrik utama yang umum digunakan dalam klasifikasi teks meliputi:

Akurasi (*Accuracy*): Proporsi total prediksi yang benar.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \dots\dots (3)$$

Presisi (*Precision*) : Kemampuan model untuk tidak melabeli hasil negatif sebagai positif. Mengukur keakuratan hasil positif yang diprediksi.

Rekal (*Recall*) : Kemampuan model untuk menemukan semua sampel positif. Mengukur kelengkapan hasil positif yang diprediksi.

Skor F1 (*F1-Score*) : Rata-rata harmonik dari Presisi dan Rekal. Metrik yang seimbang, sangat penting jika distribusi kelas tidak seimbang [21].

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad \dots\dots (4)$$

2) *Pembangunan Matriks Konfusi* : Matriks konfusi adalah alat visual dan tabular yang digunakan untuk membandingkan hasil prediksi model dengan nilai *ground truth*.

Menangani Data yang Tidak Seimbang. Dataset sentimen yang digunakan dalam penelitian ini menunjukkan distribusi yang sangat tidak seimbang, dengan ulasan positif mendominasi dataset. Untuk mengurangi bias selama evaluasi model, pembagian data bertingkat diterapkan untuk memastikan representasi kelas yang proporsional di seluruh set pelatihan, validasi, dan pengujian. Selain itu, metrik evaluasi berbobot digunakan untuk memberikan penilaian yang lebih representatif terhadap kinerja model. Meskipun pembobotan kelas dan teknik resampling tidak diterapkan dalam penelitian ini, pembagian bertingkat dan metrik evaluasi berbobot digunakan untuk mengurangi bias evaluasi. Pekerjaan di masa mendatang dapat mengeksplorasi teknik penyeimbangan tingkat data untuk meningkatkan kinerja kelas minoritas.

### III. HASIL DAN PEMBAHASAN

#### A. Karakteristik Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 1.665 ulasan pengguna aplikasi pedometer yang dikumpulkan dari Google Play Store. Setelah proses pembersihan data dan penghapusan ulasan kosong, seluruh data digunakan untuk analisis sentimen. Distribusi sentimen menunjukkan dominasi kelas positif, yang mencerminkan tingkat kepuasan pengguna terhadap aplikasi, namun juga mengindikasikan ketidakseimbangan data pada kelas netral dan negatif.

Proses pengumpulan data dilakukan secara otomatis menggunakan teknik web scraping terhadap platform *Google Playstore*. Sumber data dalam penelitian ini adalah ulasan pengguna (*user reviews*) yang terdapat pada aplikasi Pedometer di *Google Play Store*. Data yang dikumpulkan berfokus pada ulasan berbentuk teks (*textual reviews*) yang dipublikasikan dalam rentang waktu 2015 hingga 2025.

Pengambilan data ulasan dilakukan melalui pemrograman Python yang dieksekusi di lingkungan *Google Colaboratory*. Penggunaan *Colaboratory (Colab)* menjamin lingkungan komputasi yang stabil dan bebas masalah instalasi *dependency*.

Pustaka (*library*) utama yang digunakan adalah *google-play-scraper*. Pustaka ini dirancang khusus untuk mengambil data ulasan, rating, dan metadata aplikasi secara terstruktur dari *Google Playstore*.

```

colab Pedometer-IndoBERT.ipynb
--- Ringkasan Hasil Pengambilan Data ---
Jumlah Total Ulasan yang Berhasil Diambil: 1665

Lima (5) Baris Pertama dari Data Ulasan:
  reviewid  username  score  content \
0  KASTUI channel  5          bagus
1  Kizaldil 899    5          tambahkan featured tahunan rga
2  Carl Bejir 4  why does the name is Pedometer? It's should be...
3  Almar Nondiya 5          sangat membantu, terimakasih
4  Swido 62    5          sangat membantu kegiatan olga

dt
0 2025-12-08 05:11:03
1 2025-12-07 13:58:59
2 2025-12-01 15:25:35
3 2025-12-01 11:32:57
4 2025-11-30 22:57:39

Kolom dan Tipe Data:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1665 entries, 0 to 1664
Data columns (total 11 columns):
 #  Column                Non-Null Count  Dtype
---  ---                ---
0  reviewid              1665 non-null   object
1  username              1665 non-null   object
2  userimage            1665 non-null   object
3  content              1664 non-null   object
4  score                1665 non-null   float64
5  ...                  ...

```

Gambar 3. Ringkasan Hasil Pengambilan Data

**B. Hasil Pra-Pemrosesan Data**

Tahap pra-pemrosesan dilakukan untuk meningkatkan kualitas data sebelum proses pelatihan model. Proses ini meliputi *case folding*, pembersihan karakter non-alfabet, penghapusan spasi berlebih, serta pelabelan sentimen berdasarkan rating pengguna. Hasil pra-pemrosesan menghasilkan data teks yang lebih bersih dan representatif terhadap opini pengguna, sehingga dapat meningkatkan kinerja model dalam proses klasifikasi sentimen.

Selanjutnya dilakukan penanganan nilai kosong (*missing values*), khususnya pada kolom teks ulasan. Baris data yang tidak memiliki isi pada kolom content dihapus dari dataset karena tidak dapat digunakan dalam proses analisis sentimen berbasis teks. Langkah ini penting untuk menjaga konsistensi dan kualitas data, sehingga seluruh data yang digunakan benar-benar mengandung informasi tekstual yang dapat diproses oleh model.

Tahap berikutnya adalah pembersihan teks (*text cleaning*) yang bertujuan untuk menormalkan dan menyederhanakan teks ulasan. Proses ini mencakup konversi seluruh huruf menjadi huruf kecil (*case folding*), penghapusan tautan (URL), penghilangan karakter non-alfabet seperti angka dan tanda baca, serta penghapusan spasi berlebih. Hasil dari proses ini adalah teks ulasan yang lebih bersih, konsisten, dan representatif terhadap opini pengguna.

Setelah teks dibersihkan, dilakukan pelabelan sentimen berdasarkan rating bintang yang diberikan pengguna. Pelabelan ini menggunakan pendekatan heuristik sederhana, di mana rating 4 dan 5 dikategorikan sebagai sentimen positif, rating 3 sebagai sentimen netral, serta rating 1 dan 2 sebagai sentimen negatif. Pendekatan ini umum digunakan dalam penelitian analisis sentimen ulasan aplikasi karena rating numerik dianggap mencerminkan persepsi pengguna secara langsung terhadap aplikasi.

Tahap pra-pemrosesan diakhiri dengan penyajian ringkasan hasil, yang meliputi distribusi kelas sentimen, jumlah data akhir setelah pembersihan, serta contoh data yang telah dibersihkan dan diberi label. Dataset hasil pra-pemrosesan kemudian disusun ulang dengan hanya menyertakan teks ulasan bersih dan label sentimen, lalu disimpan dalam format CSV. Penyimpanan ini bertujuan untuk memastikan data siap digunakan pada tahap berikutnya,

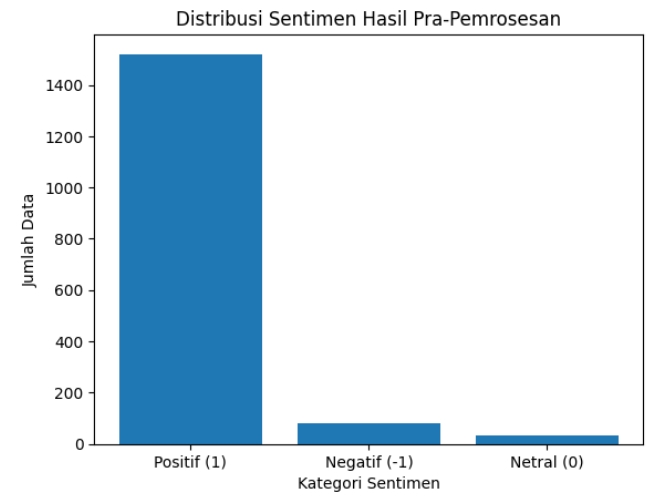
yaitu proses *fine-tuning* model IndoBERT, serta memudahkan reusabilitas data apabila diperlukan analisis lanjutan.

TABEL I  
PERBANDINGAN SETELAH PREPROCESSING

Data Asli (Raw)	Tahap Preprocessing	Hasil (Cleaned)
"Aplikasinya bagus bgt!! Tp sayang boros batre.."	<i>Case Folding &amp; Cleaning</i>	aplikasi bagus banget tapi sayang boros baterai
"Bintang 1 dulu, sering force close."	<i>Slang/ Formalization</i>	bintang satu dulu sering keluar sendiri

Analisis Kesalahan. Analisis lebih lanjut dilakukan untuk memeriksa sampel yang salah klasifikasi dalam set pengujian. Hasilnya menunjukkan bahwa ulasan negatif yang berisi keluhan ringan atau ekspresi campuran sering salah diklasifikasikan sebagai positif. Misalnya, ulasan "aplikasinya bagus tapi sering error" diprediksi sebagai positif karena adanya kata kunci positif yang dominan. Ulasan netral sangat sulit diklasifikasikan karena sering berisi pernyataan deskriptif tanpa polaritas emosional yang eksplisit. Temuan ini menunjukkan bahwa klasifikasi sentimen netral membutuhkan lebih banyak data pelatihan atau analisis sentimen berbasis aspek untuk menangkap isyarat kontekstual yang halus.

Ekspresi informal, kata-kata slang, dan kata-kata panjang yang umum ditemukan dalam ulasan Play Store dinormalisasi menggunakan aturan normalisasi manual. Emotikon dan karakter non-alfabetis dihapus karena tidak digunakan oleh tokenizer IndoBERT. Langkah ini bertujuan untuk mengurangi noise sambil mempertahankan makna semantik.



Gambar 4. Distribusi Sentimen Hasil Pra-Pemrosesan

Hasil pra-pemrosesan data menunjukkan bahwa dari total 1.665 ulasan pengguna aplikasi pedometer yang berhasil dikumpulkan, terdapat satu ulasan yang tidak memiliki konten teks sehingga dihapus pada tahap penanganan nilai kosong. Setelah proses pembersihan teks dilakukan, jumlah data akhir yang valid dan siap dianalisis menjadi sebanyak

1.636 ulasan. Proses ini memastikan bahwa seluruh data yang digunakan dalam penelitian benar-benar mengandung informasi tekstual yang relevan untuk analisis sentimen.

TABEL II  
RINGKASAN HASIL PRA-PEMROSESAN

Tahapan Pra-pemrosesan	Keterangan / Hasil
Ukuran Data Awal	1.665 baris
Pembersihan Data	Penghapusan nilai kosong ( <i>missing values</i> ) dan redundansi
Ukuran Data Akhir	1.636 baris
Distribusi Sentimen	Positif (1): 1.520 Negatif (-1): 82 Netral (0): 34
Labeling & Formatting	Konversi skor ke label sentimen dan pembersihan teks ( <i>cleaning</i> )

Distribusi sentimen pada dataset hasil pra-pemrosesan memperlihatkan ketimpangan kelas yang cukup signifikan. Kelas sentimen positif mendominasi dengan jumlah 1.520 ulasan, diikuti oleh sentimen negatif sebanyak 82 ulasan, dan sentimen netral sebanyak 34 ulasan. Pola distribusi ini mengindikasikan bahwa mayoritas pengguna memberikan respons positif terhadap aplikasi pedometer, namun sekaligus menunjukkan adanya kondisi *imbalanced* data yang berpotensi memengaruhi kinerja model klasifikasi, khususnya dalam mengenali sentimen minoritas.

Dominasi sentimen positif mencerminkan perilaku pengguna pada aplikasi kebugaran, di mana pengguna yang puas cenderung memberikan umpan balik positif yang singkat, sementara pengguna yang tidak puas memberikan ungkapan yang lebih kompleks.

Contoh lima data teratas yang ditampilkan menunjukkan bahwa teks ulasan telah berhasil dibersihkan dari karakter non-alfabet dan dinormalisasi ke dalam bentuk huruf kecil, sehingga lebih konsisten dan mudah diproses oleh model. Selain itu, setiap ulasan telah diberi label sentimen yang sesuai dengan rating bintang pengguna, yang menegaskan bahwa tahap pra-pemrosesan berhasil menghasilkan dataset bersih, terstruktur, dan siap digunakan pada tahap *fine-tuning* model IndoBERT untuk klasifikasi sentimen.

#### C. Instalasi Pustaka IndoBERT & Persiapan Dataset :

Tahap persiapan data untuk pelatihan model diawali dengan instalasi pustaka *Hugging Face* yang terdiri dari *transformers* dan *datasets*. Pustaka *transformers* digunakan sebagai kerangka utama dalam memuat dan melakukan *fine-tuning* model IndoBERT, sedangkan *datasets* dimanfaatkan untuk mengelola dataset secara efisien, terstruktur, dan kompatibel dengan proses pelatihan model berbasis *Transformer*.

Setelah lingkungan siap, data hasil pra-pemrosesan dimuat untuk tahap selanjutnya. Dataset ini sebelumnya telah disimpan dalam format terstruktur yang berisi teks ulasan bersih dan label sentimen. Selanjutnya dilakukan penyesuaian

label sentimen agar sesuai dengan kebutuhan model klasifikasi IndoBERT, yang mensyaratkan label numerik berurutan. Pada tahap ini, sentimen negatif, netral, dan positif masing-masing dipetakan ke dalam label 0, 1, dan 2. Proses ini memastikan konsistensi format label serta kompatibilitas dengan arsitektur model yang digunakan.

Dataset kemudian dipersiapkan untuk proses pelatihan dengan melakukan pembagian data menjadi tiga subset, yaitu data latih, data validasi, dan data uji dengan rasio 80:10:10. Pembagian ini dilakukan menggunakan teknik *stratified split* untuk mempertahankan proporsi kelas sentimen pada setiap subset, mengingat distribusi data yang tidak seimbang. Pendekatan ini bertujuan agar model memperoleh representasi data yang adil selama proses pelatihan dan evaluasi.

Selanjutnya, dataset dalam format *Pandas DataFrame* dikonversi ke dalam format *DatasetDict* yang disediakan oleh pustaka *Hugging Face*. Konversi ini memungkinkan integrasi dataset secara langsung dengan pipeline pelatihan model IndoBERT. Setelah konversi, dilakukan pengecekan distribusi label pada masing-masing subset data latih, validasi, dan uji untuk memastikan bahwa proporsi kelas tetap terjaga. Dengan demikian, dataset yang telah dipersiapkan dinyatakan siap untuk memasuki tahap tokenisasi dan proses *fine-tuning* model IndoBERT.

#### D. Tokenisasi Data dan Fine-Tuning IndoBERT Base

Tahap *fine-tuning* model diawali dengan pemuatan pustaka utama yang mendukung komputasi pembelajaran mendalam dan evaluasi kinerja model. Proses ini melibatkan penggunaan *PyTorch* sebagai kerangka komputasi utama, pustaka *Transformers* dari *Hugging Face* untuk memuat model dan tokenizer IndoBERT, serta pustaka *scikit-learn* untuk menghitung metrik evaluasi. Selain itu, sistem secara otomatis mendeteksi ketersediaan perangkat komputasi, baik GPU maupun CPU, untuk memastikan proses pelatihan berjalan secara optimal sesuai sumber daya yang tersedia.

Model yang digunakan dalam penelitian ini adalah IndoBERT-Base dengan varian *indobenchmark / indobert-base-pl* yang telah dilatih sebelumnya pada korpus bahasa Indonesia. Model ini dikonfigurasi untuk tugas klasifikasi teks multikelas dengan tiga label sentimen, yaitu negatif, netral, dan positif. Bersamaan dengan itu, tokenizer IndoBERT dimanfaatkan untuk mengubah teks ulasan menjadi representasi token numerik yang dapat dipahami oleh model. Seluruh input teks dipotong atau dipadatkan hingga panjang maksimum 128 token untuk menjaga konsistensi dimensi masukan selama pelatihan.

Proses tokenisasi diterapkan pada seluruh dataset yang telah dibagi sebelumnya ke dalam data latih, validasi, dan uji. Tokenisasi dilakukan secara *batch processing* dengan mekanisme *padding* dan *truncation* agar setiap sampel memiliki panjang token yang seragam. Tahap ini merupakan prasyarat penting sebelum data dapat diproses oleh arsitektur *Transformer* dalam model IndoBERT.

Untuk mengevaluasi kinerja model secara kuantitatif, digunakan sejumlah metrik evaluasi standar, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Perhitungan metrik dilakukan menggunakan pendekatan *weighted average* menyesuaikan kondisi distribusi data yang tidak seimbang. Pendekatan ini memungkinkan evaluasi yang lebih representatif terhadap performa model pada seluruh kelas sentimen.

Proses pelatihan model dilakukan dengan menetapkan sejumlah *hyperparameter* utama, meliputi ukuran batch sebesar 16, *learning rate* sebesar  $2 \times 10^{-5}$ , serta jumlah epoch sebanyak tiga kali pelatihan. Konfigurasi ini dipilih untuk menjaga keseimbangan antara performa model dan efisiensi komputasi. Selain itu, diterapkan mekanisme *early model selection* dengan menyimpan model terbaik berdasarkan nilai *weighted F1-score* pada data validasi, sehingga model yang dihasilkan memiliki kemampuan generalisasi yang optimal.

Tahap *fine-tuning* dijalankan menggunakan modul *Trainer* dari pustaka *Hugging Face* yang mengintegrasikan model, dataset, parameter pelatihan, serta fungsi evaluasi ke dalam satu pipeline terstruktur. Setelah proses pelatihan selesai, model terbaik beserta *tokenizer* disimpan untuk digunakan pada tahap evaluasi akhir dan analisis lebih lanjut. Hasil *fine-tuning* menunjukkan bahwa model IndoBERT-Base mampu mempelajari pola sentimen pada ulasan aplikasi pedometer secara efektif, sebagaimana ditunjukkan pada hasil pelatihan yang disajikan pada Gambar 5.

```

colab Pedometer-IndoBERT.ipynb
tokenizer.save_pretrained(final_model_path)
print(f"[Langkah 4 Selesai] Model terbaik telah disimpan di: {final_model_path}")

Panggilan perintah: cuda
-- Memuat tokenizer dan Model IndoBERT Base (IndoBERT-base-pt) --
Some weights of BertForSequenceClassification were not initialized from the model checkpoint at indochecker/indoBERT-base-pt and are newly initialized: ['classifier', 'dropout']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Pre: 0E: [ 0.1388 [0.0000], 1 examples/1]
Pre: 0E: [ 0.1384 [0.0000], 1 examples/1]
Pre: 0E: [ 0.1384 [0.0000], 1 examples/1]
FileNotFoundError: [Errno 2] No such file or directory: 'tokenizer' is deprecated and will be removed in version 5.0.0 for "Trainer". Use "processing_class" in "DataCollatorWithTokenizer" instead.

wandb: (1) Create a W&B account
wandb: (2) Use an existing W&B account
wandb: (3) Don't visualize my results
wandb: Enter your choice
-- Minimal Process Fine-Tuning IndoBERT Base --
wandb: You chose "Don't visualize my results"
Tracking on with wandb offline. In this directory, run "wandb online" or set WANDB_MODE=online to enable cloud syncing.
Run data to saved locally in /root/.wandb/offline-run-20231019_131120-getfich7
-----
Epoch Training Loss Validation Loss Accuracy F1 weighted Precision weighted Recall weighted
1 0.716500 0.227107 0.932927 0.903404 0.875602 0.932927
2 0.682000 0.173337 0.951920 0.947034 0.846637 0.951920
3 0.108700 0.229802 0.957917 0.948319 0.904493 0.957917
-----
[Langkah 4 Selesai] Model terbaik telah disimpan di: ./results_indoBERT/pedometer/final_indoBERT_pedometer
    
```

Gambar 5. Hasil Proses *Fine-Tuning IndoBERT Base*

E. Evaluasi Akhir Model pada Test Set

Tahap evaluasi model dilakukan menggunakan data uji (test set) yang sebelumnya telah dipisahkan dari data latih dan data validasi. Evaluasi ini bertujuan untuk mengukur kemampuan generalisasi model IndoBERT-Base *fine-tuning* dalam mengklasifikasikan sentimen ulasan yang belum pernah dilihat selama proses pelatihan. Proses evaluasi diawali dengan penerapan fungsi evaluasi bawaan pada model terlatih untuk memperoleh nilai metrik kinerja secara keseluruhan.

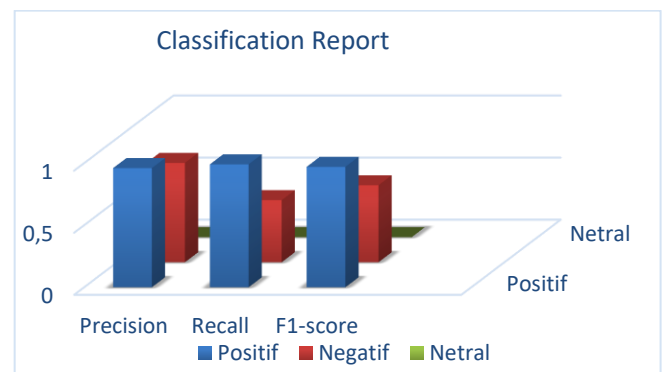
Selain evaluasi global, dilakukan pula proses prediksi pada seluruh data uji untuk analisis kinerja yang lebih mendalam. Hasil prediksi model kemudian dibandingkan dengan label sebenarnya guna menghasilkan laporan klasifikasi (*classification report*) yang menyajikan nilai

*precision*, *recall*, dan *F1-score* untuk setiap kelas sentimen, yaitu negatif, netral, dan positif. Pendekatan ini memungkinkan analisis yang lebih terperinci terhadap performa model pada masing-masing kelas, khususnya pada kelas minoritas yang cenderung sulit dikenali.

TABEL III  
PERFORMA KESELURUHAN MODEL (EVALUASI FINAL)

Metrik Evaluasi	Nilai (Desimal)	(Persentase)
Akurasi (Accuracy)	0,9451	94,51%
F1-Score (Weighted)	0,9329	93,29%
Precision (Weighted)	0,9247	92,47%
Loss	0,2671	-

Untuk memperkuat interpretasi hasil, evaluasi juga dilengkapi dengan matriks kebingungan (*confusion matrix*). Matriks ini menggambarkan distribusi prediksi benar dan salah pada setiap kelas sentimen, sehingga dapat diketahui pola kesalahan klasifikasi yang dilakukan oleh model. Melalui matriks ini, dapat dianalisis kecenderungan model dalam mengklasifikasikan ulasan ke kelas tertentu serta mengidentifikasi kelas sentimen yang paling sering tertukar.



Gambar 6. Diagram *Classification Report*

Berdasarkan hasil evaluasi pada data uji, model IndoBERT-Base *fine-tuning* mencapai tingkat akurasi akhir sebesar 94,51%. Meskipun nilai akurasi tergolong sangat tinggi, analisis lebih lanjut menunjukkan bahwa tingginya akurasi tersebut dipengaruhi oleh dominasi kelas sentimen positif dalam dataset. Oleh karena itu, metrik F1-score, khususnya pada kelas negatif dan netral, menjadi indikator yang lebih representatif dalam menilai kinerja model secara menyeluruh. Temuan ini menegaskan bahwa meskipun model mampu bekerja dengan sangat baik secara umum, tantangan utama masih terletak pada penanganan ketidakseimbangan data sentimen.

TABEL IV  
LAPORAN KLASIFIKASI RINCI

Label Sentimen	Precision	Recall	F1-Score	Support
Negatif (0)	0,80	0,50	0,62	8
Netral (1)	0,00	0,00	0,00	4
Positif (2)	0,96	0,99	0,97	152
Macro Avg	0,59	0,50	0,53	164
Weighted Avg	0,92	0,95	0,93	164



Akurasi Akhir Model (Evaluasi pada Test Set) menunjukkan 94.51%. Adapun target akurasi 100% pada data sangat tidak seimbang (*Imbalanced Data*) rentan *Overfitting*.

Model ini menunjukkan kinerja yang sangat baik secara keseluruhan, meskipun ada beberapa catatan penting karena ketidakseimbangan data.

Meskipun akurasi global tinggi, performa untuk setiap kelas menunjukkan dampak signifikan dari ketidakseimbangan data (*Imbalanced Data*).

Akurasi tinggi (94.51%) didorong oleh kinerja luar biasa pada kelas Positif (2) yang dominan (152 dari 164 sampel). Ini adalah alasan mengapa akurasinya mendekati 100%.

Hasil interpretasi matriks kebingungan menunjukkan bahwa model *IndoBERT-Base fine-tuning* memiliki kinerja yang sangat baik dalam mengklasifikasikan sentimen positif. Dari total 152 ulasan yang berlabel positif, sebanyak 151 ulasan berhasil diprediksi dengan benar, sementara hanya satu ulasan yang mengalami kesalahan klasifikasi dan diprediksi sebagai sentimen netral. Temuan ini mengindikasikan bahwa model mampu mengenali pola linguistik pada ulasan positif secara konsisten dan akurat.

TABEL V  
Matriks Kebingungan (*Confusion Matrix*)

Label Sebenarnya ↓ / Label Prediksi →	Negatif (0)	Netral (1)	Positif (2)
Negatif (0)	4 (Benar)	0	4 (Salah prediksi sebagai Positif)
Netral (1)	1	0 (Salah)	3 (Salah prediksi sebagai Positif)
Positif (2)	0	1	151 (Benar)

Pada kelas sentimen negatif, kinerja model menunjukkan hasil yang cukup moderat. Dari delapan ulasan negatif yang terdapat pada data uji, sebanyak empat ulasan berhasil diklasifikasikan dengan benar, sedangkan empat ulasan lainnya salah diprediksi sebagai sentimen positif. Hal ini mengindikasikan adanya tumpang tindih karakteristik bahasa antara ulasan negatif dan positif, serta menunjukkan keterbatasan model dalam membedakan kedua kelas tersebut, terutama ketika jumlah data negatif relatif sedikit.

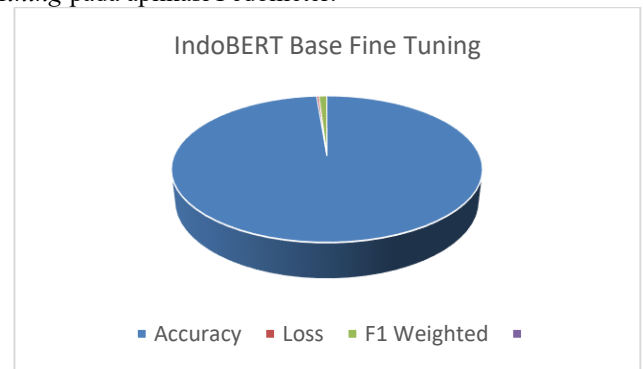
Sementara itu, kinerja model pada kelas sentimen netral menunjukkan hasil yang kurang optimal. Dari empat ulasan netral yang diuji, tidak satu pun berhasil diprediksi dengan benar. Ulasan netral tersebut masing-masing salah diklasifikasikan ke dalam kelas negatif sebanyak satu ulasan dan kelas positif sebanyak tiga ulasan. Kondisi ini menegaskan bahwa kelas netral merupakan kelas yang paling sulit dikenali oleh model, yang kemungkinan besar disebabkan oleh jumlah data yang sangat terbatas serta karakteristik bahasa yang cenderung ambigu dan berada di antara sentimen positif dan negatif.

Akurasi keseluruhan yang tinggi yang dicapai oleh model ini sebagian besar dipengaruhi oleh dominasi ulasan positif dalam dataset. Hal ini mencerminkan perilaku pengguna di

dunia nyata, di mana pengguna yang puas lebih cenderung memberikan umpan balik positif yang singkat, sementara ulasan negatif cenderung lebih sedikit tetapi secara linguistik lebih kompleks. Akibatnya, meskipun *IndoBERT-Base* secara efektif menangkap pola sentimen positif, kinerjanya pada kelas netral dan negatif tetap terbatas karena kelangkaan data dan ambiguitas linguistik.

#### IV. KESIMPULAN

Model *IndoBERT Base Fine-Tuning* telah menghasilkan akurasi yang sangat tinggi (94.51%) untuk tugas *Opinion Mining* pada aplikasi Pedometer.



Gambar 8. Diagram Accuracy IndoBERT-Base

Meskipun Akurasi mencapai 94.51%, analisis mendalam melalui *Classification Report* dan *Confusion Matrix* mengungkapkan bahwa kinerja model sangat dipengaruhi oleh ketidakseimbangan data. Model menunjukkan performa Superior pada kelas Positif (*F1-score* 0.97), namun lemah dalam mengidentifikasi kelas Netral dan memiliki Recall rendah (0.50) pada kelas Negatif. Hal ini mengindikasikan bahwa model cenderung memprediksi kelas dominan (Positif).

#### UCAPAN TERIMA KASIH

Penulis menyampaikan penghargaan dan terima kasih yang sebesar-besarnya atas bantuan dan dukungan yang telah diberikan selama pelaksanaan penelitian ini. Rasa terima kasih yang mendalam disampaikan secara khusus kepada Yayasan Sosial Bangun Putra Bangsa atas dukungan finansial yang krusial, yang telah memungkinkan terlaksananya seluruh tahapan riset ini dari awal hingga akhir.

#### DAFTAR PUSTAKA

- [1] B. Lin, N. Cassee, A. Serebrenik, G. Bavota, N. Novielli, and M. Lanza, "Opinion Mining for Software Development: A Systematic Literature Review," *ACM Trans. Softw. Eng. Methodol.*, vol. 31, no. 3, 2022, doi: 10.1145/3490388.
- [2] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, "Sentiment analysis and opinion mining on educational data: A survey," *Nat. Lang. Process. J.*, vol. 2, no. October 2022, p. 100003, 2023, doi: 10.1016/j.nlp.2022.100003.

- [3] S. Biswas and G. Poornalatha, "Opinion Mining Using Multi-Dimensional Analysis," *IEEE Access*, vol. 11, no. March, pp. 25906–25916, 2023, doi: 10.1109/ACCESS.2023.3256521.
- [4] G. R. McCormack, J. C. Spence, T. L. McHugh, and W. Kerry Mummery, "The effect of neighborhood walkability on changes in physical activity and sedentary behavior during a 12-week pedometer-facilitated intervention," *PLoS One*, vol. 17, no. 12 December, 2022, doi: 10.1371/journal.pone.0278596.
- [5] F. Baharuddin, "Fine-Tuning IndoBERT for Indonesian Exam Question Classification Based on Bloom 's Taxonomy," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 9, no. 2.
- [6] R. I. Perwira and V. A. Permadi, "Domain-Specific Fine-Tuning of IndoBERT for Aspect- Based Sentiment Analysis in Indonesian Travel User- Generated Content," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 11, no. 1, pp. 30–40, 2025.
- [7] A. Simanjuntak, R. Lumbantoruan, K. Sianipar, R. Gultom, M. Simaremare, and S. Situmeang, "Research and Analysis of IndoBERT Hyperparameter Tuning in Fake News Detection," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 13, no. 1, pp. 60–67, 2024, doi: 10.22146/jnteti.v13i1.8532.
- [8] D. Nuryadi et al., "Fine Tuning Indobert Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Tiket.Com Di Google Play Store," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 3577–3583, 2025, doi: 10.36040/jati.v9i2.13204.
- [9] R. A. E. V. T. Sapanji, D. Hamdani, and P. Harahap, "Sentiment Analysis of the Top 5 E-commerce Platforms in Indonesia using Text Mining and Natural Language Processing (NLP)," *J. Appl. Informatics Comput.*, vol. 7, no. 2, pp. 202–211, 2023, doi: 10.30871/jaic.v7i2.6517.
- [10] F. Fachrurrazi, "Perancangan Aplikasi Penghitung Langkah Kaki Berbasis Android Menggunakan MIT App Inventor," *J. Sains dan Teknol.*, vol. 01, no. 01, pp. 19–25, 2024.
- [11] G. Daruhadi and P. Sopiati, "Metode Pengumpulan Data Penelitian," *J-CEKI J. Cendekia Ilm.*, vol. 3, no. 5, pp. 5423–5443, 2024.
- [12] A. N. A. Zulkifli, and N. A. Shafie, "Review of the Lazada application on Google Play Store: sentiment analysis," *J. Comput. Res. Innov.*, vol. 9, no. 1, 2024, doi: 10.24191/jcrinn.v9i1.412.
- [13] M. F. Kono, I. N. Fajri, and Y. Pristyanto, "Public Sentiment Analysis on Corruption Issues in Indonesia Using IndoBERT Fine-Tuning, Logistic Regression, and Linear SVM," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2616–2628, 2025, doi: 10.30871/jaic.v9i5.10537.
- [14] A. ' Zhahrina, U. Sofiah, D. Wahyu, A. Andayani, N. Khairurrabbani, and F. Saputri, "Penerapan Metode Support Vector Machine (SVM) dalam Analisis Sentimen Ulasan Aplikasi Tokopedia di Google Play Store," *SEMINAR NASIONAL CORISINDO* no. September 2025, pp. 84–91.
- [15] A. Prabowo and F. Indra Sanjaya, "Penerapan Metode Transfer Learning Pada Indobert Untuk Analisis Sentimen Teks Bahasa Jawa Ngoko Lugu," *Simkom*, vol. 9, no. 2, pp. 205–217, 2024, doi: 10.51717/simkom.v9i2.478.
- [16] D. Y. A. Ningrum, E. Daniati, and M. N. Muzaki, "Perbandingan Model BERT dan RNN-LSTM pada Analisis Sentimen Aplikasi BRI Mobile," *IJCSR Indones. J. Comput. Sci. Res.*, vol. 4, no. 2, pp. 75–85, 2025, [Online]. Available: <https://subset.id/index.php/IJCSR>
- [17] R. Rahmadani, A. Rahim, and R. Rudiman, "Analisis Sentimen Ulasan 'Ojol the Game' Di Google Play Store Menggunakan Algoritma Naive Bayes Dan Model Ekstraksi Fitur Tf-Idf Untuk Meningkatkan Kualitas Game," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4988.
- [18] R. Ulgasesa, A. B. P. Negara, and T. Tursina, "Pengaruh Stemming Terhadap Performa Klasifikasi Sentimen Masyarakat Tentang Kebijakan New Normal," *J. Sist. dan Teknol. Inf.*, vol. 10, no. 3, p. 286, 2022, doi: 10.26418/justin.v10i3.53880.
- [19] Z. G. Ramadhan, W. M. Ashari, and M. Kopravi, "Perbandingan Kinerja Algoritma Machine Learning Deteksi Malware dengan Z-Score Normalization Hasil Terbaik pada Random Forest," *Jurnal Algoritma*. pp. 1670–1681, 2025, doi: 10.33364/algoritma/v.22-2.3077.
- [20] T. Aprilah, D. R. Ignatius, M. Setiadi, and W. Herowati, "Enhancing Aspect-Based Sentiment Analysis via Hugging Face Fine-Tuned IndoBERT," *J. Appl. Informatics Comput.*, vol. 9, no. 6, pp. 3821–3830, 2025.
- [21] Muhammad Bayu Nugroho, Akhmad Khanif Zyen, and Nur Aeni Widiastuti, "Multiclass Sentiment Analysis of Electric Vehicle Incentive Policies Using IndoBERT and DeBERTa Algorithms," *J. Appl. Informatics Comput.*, vol. 9, no. 3, pp. 910–919, 2025, doi: 10.30871/jaic.v9i3.9511.