E-ISSN: 2964-6219, P-ISSN: 2964-6200
Volume 4 Number 2 August 2026.
pp: 224-237. DOI: https://doi.org/10.58477/cj.v4i2.497.

Computer Journal

RESEARCH ARTICLE Open Access

Analisis Sentimen Ulasan Pengguna inDrive Menggunakan IndoBERT dan Algoritma Genetika pada Klasifikasi *K-Nearest Neighbor*

Muhammad Sigit Nurhafid [1*], **Rudiman** [2], **Taghfirul Azhima Yoga** [3]

[1*,2,3] Jurusan S1 Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur, Kota Samarinda, Provinsi Kalimantan Timur, Indonesia.

*Correspondence email: 2011102441077@umkt.ac.id.

Received: 21 August 2026.
Revised: 25 August 2026.
Accepted: 29 August 2026.
Published: 30 August 2026.

Full list of author information is available at the end of the article.

Abstract

This study analyzes sentiment in InDrive user reviews from the Google Play Store using IndoBERT, Genetic Algorithm (GA), and K-Nearest Neighbor (KNN). A total of 2,000 reviews were assigned to three sentiment classes: 1,695 negative, 235 positive, and 70 neutral reviews. The pretrained indobenchmark/indobert-base-p1 model was used as a feature extractor by taking the [CLS] representation to produce 768-dimensional embeddings. The dataset was divided using a stratified 80:20 split into 1,600 training and 400 testing samples. The optimal K value was determined through stratified five-fold cross-validation on the training data. GA was applied only to the training set using a population of 30 individuals, 25 generations, a crossover rate of 0.8, a mutation rate of 0.005, and a feature penalty of 0.002. GA selected 250 features, reducing the dimensionality by 67.45%. IndoBERT + KNN correctly classified 365 of 400 test samples, achieving 91.25% accuracy (95% CI: 88.07–93.64%) and a macro F1-score of 66.88%. IndoBERT + GA + KNN correctly classified 362 samples, achieving 90.50% accuracy (95% CI: 87.23–93.00%) and a macro F1-score of 62.71%. Both models exceeded the 84.75% majority-class baseline. However, only three and two of the 14 neutral samples were correctly classified, respectively. GA substantially reduced feature dimensionality but did not improve predictive performance, indicating a trade-off between representation compactness and minority-class classification performance.

Keywords: Sentiment Analysis; IndoBERT; Genetic Algorithm; K-Nearest Neighbor.

Abstrak

Penelitian ini menganalisis sentimen ulasan pengguna InDrive pada Google Play Store menggunakan IndoBERT, Algoritma Genetika (GA), dan K-Nearest Neighbor (KNN). Sebanyak 2.000 ulasan dikategorikan ke dalam tiga kelas sentimen, yaitu 1.695 ulasan negatif, 235 positif, dan 70 netral. Model pretrained indobenchmark/indobert-base-p1 digunakan sebagai feature extractor dengan mengambil representasi token [CLS] sehingga menghasilkan embedding berdimensi 768. Dataset dibagi secara stratified dengan rasio 80:20 menjadi 1.600 data latih dan 400 data uji. Nilai K ditentukan melalui Stratified 5-Fold Cross-Validation pada data latih. GA dijalankan hanya pada data latih dengan populasi 30 individu, 25 generasi, crossover rate 0,8, mutation rate 0,005, dan penalti fitur 0,002. GA memilih 250 fitur sehingga mereduksi dimensi sebesar 67,45%. IndoBERT + KNN menghasilkan 365 prediksi benar dari 400 data uji dengan akurasi 91,25% (95% CI: 88,07–93,64%) dan macro F1-score 66,88%. IndoBERT + GA + KNN menghasilkan 362 prediksi benar dengan akurasi 90,50% (95% CI: 87,23–93,00%) dan macro F1-score 62,71%. Kedua model melampaui majority-class baseline sebesar 84,75%. Namun, hanya tiga dari 14 sampel netral yang diklasifikasikan dengan benar oleh IndoBERT + KNN dan dua sampel oleh IndoBERT + GA + KNN. GA mampu mereduksi dimensi fitur secara substansial, tetapi belum meningkatkan performa prediktif, yang menunjukkan adanya kompromi antara keringkasan representasi dan kemampuan klasifikasi kelas minoritas.

Kata Kunci: Analisis Sentimen; IndoBERT; Algoritma Genetika; K-Nearest Neighbor.

1. Pendahuluan

Perkembangan layanan transportasi berbasis aplikasi telah membentuk pola interaksi baru antara pengguna dan penyedia layanan. Pengguna tidak hanya memanfaatkan aplikasi untuk memperoleh layanan transportasi, tetapi juga menyampaikan pengalaman mereka melalui ulasan pada toko aplikasi. Ulasan tersebut dapat memuat apresiasi, kritik, keluhan teknis, penilaian terhadap pelayanan, maupun pengalaman selama menggunakan aplikasi. Banyaknya data ulasan berbentuk teks menyebabkan proses analisis secara manual membutuhkan waktu dan konsistensi penilaian, sehingga diperlukan pendekatan komputasional untuk mengidentifikasi kecenderungan sentimen secara otomatis. Analisis sentimen merupakan salah satu bidang Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi kecenderungan opini atau sikap yang terkandung dalam teks. Pada konteks ulasan aplikasi, sentimen dapat diklasifikasikan ke dalam kategori positif, netral, dan negatif. Pemanfaatan ulasan Google Play Store sebagai sumber data analisis sentimen telah diterapkan pada berbagai layanan digital di Indonesia. Pradhisa dan Fajriyah (2024) menerapkan IndoBERT pada ulasan aplikasi e-commerce, sedangkan Tarwoto *et al*. (2025) menggunakan pendekatan serupa pada ulasan Mobile JKN. Penelitian pada aplikasi PLN Mobile, BRImo, Shopee, dan Access by KAI juga menunjukkan bahwa ulasan pengguna dapat diolah untuk mengidentifikasi pola sentimen terhadap layanan digital (Aras *et al*., 2024; Asri *et al*., 2025; Simarmata & Sasongko, 2025; Wirayudha *et al*., 2025). Selain itu, Brando *et al*. (2025) menerapkan model berbasis Transformer, yaitu BERT dan RoBERTa, untuk menganalisis sentimen ulasan pengguna aplikasi Info BMKG pada Google Play Store. Karakteristik bahasa pada ulasan pengguna menimbulkan tantangan karena teks dapat mengandung variasi struktur kalimat, bahasa informal, singkatan, serta konteks yang tidak selalu dapat direpresentasikan melalui pencocokan kata secara sederhana. Model Bidirectional Encoder Representations from Transformers (BERT) mampu menghasilkan representasi kontekstual dengan mempertimbangkan hubungan antarkata dalam teks. Dalam konteks bahasa Indonesia, IndoBERT telah digunakan pada berbagai tugas klasifikasi teks, termasuk analisis sentimen produk Tokopedia, ulasan Shopee, opini mengenai telepon pintar, dan klasifikasi teks berbahasa Indonesia lainnya (Aras *et al*., 2024; Nabiilah *et al*., 2024; Supriyadi & Sibaroni, 2023; Wau, 2025). Talaat (2023) juga mengembangkan sistem klasifikasi analisis sentimen menggunakan model hybrid berbasis BERT.

Penelitian ini menggunakan model pretrained indobenchmark/indobert-base-p1 sebagai feature extractor untuk membentuk representasi numerik setiap ulasan. Berdasarkan konfigurasi eksperimen, setiap teks direpresentasikan dalam vektor berdimensi 768. Meskipun representasi berdimensi tinggi dapat menyimpan informasi yang kaya, tidak seluruh dimensi memiliki kontribusi yang sama terhadap proses klasifikasi. Kondisi tersebut membuka peluang penerapan seleksi fitur untuk memperoleh subset fitur yang lebih ringkas tanpa mengabaikan kemampuan klasifikasi. Algoritma Genetika (Genetic Algorithm/GA) merupakan metode pencarian berbasis mekanisme evolusi yang dapat digunakan untuk menentukan kombinasi fitur dari ruang pencarian yang besar. Namun, seleksi fitur menggunakan GA tidak selalu menghasilkan peningkatan performa classifier. Putra *et al*. (2022), misalnya, menunjukkan bahwa penggunaan seleksi fitur berbasis GA belum menghasilkan akurasi yang lebih tinggi dibandingkan penggunaan keseluruhan fitur pada eksperimen analisis sentimen yang dilakukan. Penelitian lain menunjukkan bahwa keberhasilan seleksi fitur perlu dinilai dengan mempertimbangkan keseimbangan antara jumlah fitur yang dipertahankan dan performa model setelah seleksi (Mostafa *et al*., 2023; Sağbaş, 2024). Hal tersebut menunjukkan perlunya pengujian empiris untuk mengetahui apakah reduksi fitur pada representasi IndoBERT dapat menghasilkan representasi yang lebih ringkas dengan performa klasifikasi yang tetap memadai. K-Nearest Neighbor (KNN) dapat digunakan untuk mengklasifikasikan representasi numerik teks berdasarkan kedekatan antarvektor. Gaol dan Ichwani (2025) membandingkan penggunaan IndoBERT dan KNN pada analisis sentimen ulasan Tokopedia dan menunjukkan adanya perbedaan kemampuan antara pendekatan berbasis transformer dan pendekatan berbasis tetangga terdekat. Dalam penelitian ini, IndoBERT tidak digunakan sebagai classifier akhir, melainkan sebagai pembentuk representasi fitur yang selanjutnya diklasifikasikan menggunakan KNN. Dengan demikian, pengaruh seleksi fitur GA terhadap classifier KNN dapat diuji secara langsung melalui perbandingan antara penggunaan seluruh fitur IndoBERT dan subset fitur hasil seleksi GA.

Aspek lain yang perlu diperhatikan adalah ketidakseimbangan jumlah sampel antarkelas. Dataset penelitian terdiri atas 1.695 ulasan negatif, 235 ulasan positif, dan 70 ulasan netral, sehingga kelas negatif jauh lebih dominan dibandingkan kelas lainnya. Pada data multiclass yang tidak seimbang, akurasi dapat dipengaruhi oleh keberhasilan model dalam mengenali kelas mayoritas. Oleh karena itu, pemilihan metrik evaluasi menjadi penting untuk memberikan gambaran performa yang lebih representatif pada setiap kelas. Riyanto *et al*. (2023) menunjukkan pentingnya pemilihan metrik evaluasi pada klasifikasi teks multiclass

yang tidak seimbang. Penelitian Habbat *et al*. (2023), Cahya *et al*. (2024), serta Sani dan Sarwani (2024) juga menunjukkan pentingnya memperhatikan performa kelas minoritas pada model berbasis BERT dan klasifikasi teks. Berdasarkan permasalahan tersebut, penelitian ini bertujuan menganalisis sentimen ulasan pengguna InDrive menggunakan representasi IndoBERT, Algoritma Genetika sebagai metode seleksi fitur, dan K-Nearest Neighbor sebagai classifier. Pengujian dilakukan melalui dua skenario, yaitu KNN menggunakan seluruh 768 fitur IndoBERT dan KNN menggunakan 250 fitur yang dipilih oleh GA. Kedua skenario dibandingkan berdasarkan akurasi, macro precision, macro recall, macro F1-score, weighted F1-score, serta tingkat reduksi fitur yang diperoleh. Kontribusi penelitian ini diarahkan pada pengujian empiris penerapan Algoritma Genetika terhadap representasi berdimensi tinggi yang dihasilkan oleh IndoBERT. Keberhasilan GA tidak hanya dinilai berdasarkan kemungkinan peningkatan akurasi, tetapi juga berdasarkan kemampuannya memperoleh subset fitur yang lebih kecil serta perubahan performa klasifikasi yang dihasilkan. Dengan pendekatan tersebut, penelitian ini memberikan gambaran mengenai trade-off antara pengurangan dimensi fitur dan kemampuan klasifikasi pada ulasan pengguna InDrive yang memiliki distribusi kelas tidak seimbang.

2. Tinjauan Pustaka

Analisis sentimen merupakan bagian dari *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi kecenderungan opini atau sikap yang terkandung dalam teks ke dalam kategori tertentu, seperti positif, netral, dan negatif. Pada ulasan aplikasi, analisis sentimen dapat digunakan untuk menggambarkan respons pengguna terhadap layanan berdasarkan pengalaman yang disampaikan melalui komentar. Ulasan pada Google Play Store banyak dimanfaatkan sebagai sumber data analisis sentimen karena menyediakan informasi yang berasal langsung dari pengalaman pengguna terhadap suatu aplikasi (Tarwoto *et al*., 2025). Dalam penelitian ini, analisis sentimen digunakan untuk mengidentifikasi kecenderungan sentimen pada ulasan pengguna InDrive berdasarkan tiga kategori sentimen. IndoBERT merupakan model bahasa berbasis arsitektur *Bidirectional Encoder Representations from Transformers* (BERT) yang dikembangkan untuk merepresentasikan teks berbahasa Indonesia secara kontekstual. Arsitektur tersebut memungkinkan model mempertimbangkan hubungan antarkata dalam suatu urutan teks sehingga representasi yang dihasilkan tidak hanya bergantung pada kemunculan kata secara individual. Penelitian ini menggunakan model *pretrained* indobenchmark/indobert-base-p1 sebagai *feature extractor*. Representasi token [CLS] digunakan untuk merepresentasikan setiap teks dalam bentuk vektor berdimensi 768. IndoBERT telah diterapkan dalam berbagai penelitian klasifikasi sentimen pada teks berbahasa Indonesia dan menunjukkan kemampuan dalam merepresentasikan karakteristik bahasa pada ulasan pengguna (Pradhisa & Fajriyah, 2024). *K-Nearest Neighbor* (KNN) merupakan algoritma *supervised learning* yang menentukan kelas suatu data berdasarkan kelas sejumlah data latih yang memiliki kedekatan paling tinggi terhadap data tersebut. Nilai K menunjukkan jumlah tetangga terdekat yang digunakan dalam proses klasifikasi. Kedekatan antardata ditentukan berdasarkan ukuran jarak pada ruang fitur sehingga KNN dapat diterapkan pada representasi numerik teks. Dalam penelitian ini, vektor embedding yang dihasilkan oleh IndoBERT digunakan sebagai ruang fitur untuk menentukan kedekatan antarulasan sebelum proses klasifikasi. Penggunaan KNN dalam klasifikasi sentimen pada teks berbahasa Indonesia juga telah diterapkan pada penelitian terdahulu (Gaol & Ichwani, 2025).

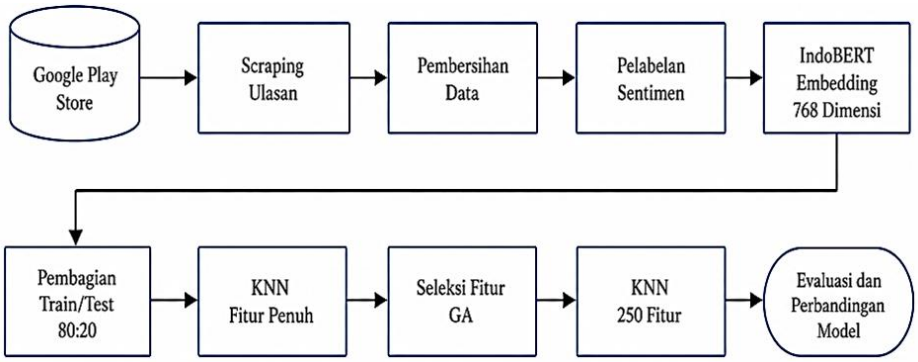
Algoritma Genetika (*Genetic Algorithm*/GA) merupakan metode optimasi berbasis mekanisme evolusi yang menggunakan populasi kandidat solusi dan proses seleksi untuk memperoleh solusi yang semakin baik berdasarkan fungsi *fitness*. Proses utama GA meliputi pembentukan populasi, seleksi, *crossover*, mutasi, dan evaluasi *fitness*. Dalam konteks seleksi fitur, setiap kandidat solusi dapat merepresentasikan kombinasi fitur yang dipertahankan atau dihilangkan. GA kemudian digunakan untuk mencari subset fitur yang memberikan keseimbangan antara jumlah fitur dan performa klasifikasi. Keberhasilan seleksi fitur tidak selalu ditunjukkan oleh peningkatan akurasi karena pengurangan jumlah fitur dapat menghasilkan kompromi antara representasi fitur dan performa classifier (Sağbaş, 2024; Tarwoto *et al*., 2025). Dalam penelitian ini, GA digunakan untuk mencari subset dari 768 fitur hasil representasi IndoBERT yang selanjutnya digunakan sebagai input bagi KNN. Data tidak seimbang terjadi ketika jumlah sampel antarkelas berbeda secara signifikan. Kondisi tersebut dapat menyebabkan classifier lebih banyak mempelajari pola kelas mayoritas sehingga akurasi yang tinggi belum tentu mencerminkan kemampuan model dalam mengenali seluruh kelas. Oleh karena itu, evaluasi pada data tidak seimbang perlu mempertimbangkan metrik yang memberikan perhatian lebih seimbang terhadap setiap kelas, seperti *macro precision*, *macro recall*, dan *macro F1-score*, selain akurasi dan *weighted F1-score* (Riyanto *et al*., 2023). Penelitian berbasis BERT juga menunjukkan bahwa ketidakseimbangan kelas dapat memengaruhi performa pada kelas minoritas. Berbagai pendekatan, seperti SMOTE, *random oversampling*, augmentasi data, dan *class*

weighting, dapat digunakan untuk menangani permasalahan ketidakseimbangan kelas pada kondisi tertentu (Cahya *et al.*, 2024; Habbat *et al.*, 2023). Dalam penelitian ini, kondisi ketidakseimbangan kelas dipertimbangkan terutama dalam pemilihan metrik evaluasi, sedangkan proses klasifikasi dilakukan tanpa penerapan teknik penyeimbangan kelas.

3. Metode

3.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk melakukan klasifikasi sentimen multiclass. Tahapan penelitian meliputi pengumpulan ulasan pengguna InDrive dari Google Play Store, pembentukan dataset, pra-pemrosesan teks, pembentukan representasi teks menggunakan IndoBERT, pembagian dataset menjadi data latih dan data uji, klasifikasi menggunakan KNN dengan seluruh fitur IndoBERT, seleksi fitur menggunakan Algoritma Genetika, klasifikasi menggunakan subset fitur hasil seleksi GA, dan evaluasi performa kedua skenario. Alur penelitian ditunjukkan pada Gambar 1.



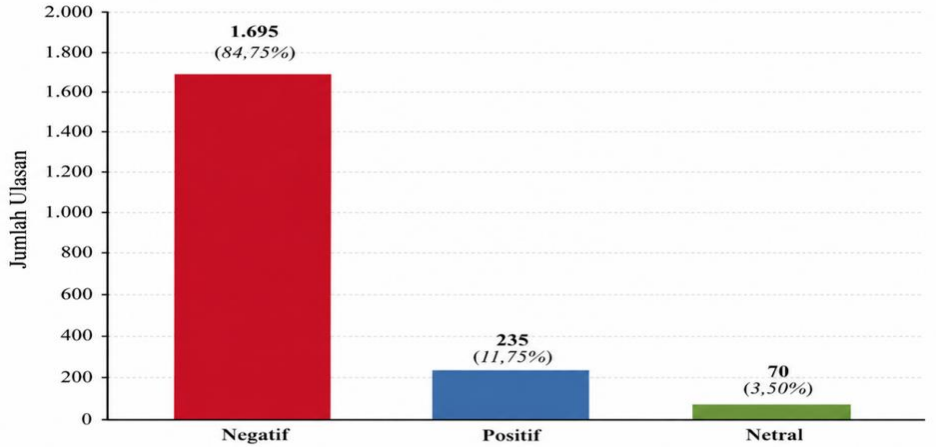
Gambar 1. Alur penelitian analisis sentimen ulasan pengguna InDrive.

3.2 Pengumpulan Data

Data penelitian berasal dari ulasan pengguna aplikasi InDrive pada Google Play Store yang dikumpulkan menggunakan bahasa pemrograman Python pada lingkungan Google Colab dengan library google-play-scraper. Pengambilan data dilakukan menggunakan ID aplikasi sinet.startup.inDriver dengan parameter bahasa lang='id' dan negara country='id'. Ulasan diambil menggunakan Sort.MOST_RELEVANT dengan count=2000, sedangkan parameter filter_score_with=None digunakan agar ulasan dari seluruh tingkat rating dapat diperoleh tanpa pembatasan skor. Proses scraping menghasilkan 2.000 ulasan dengan atribut yang digunakan meliputi nama pengguna (userName), rating (score), waktu publikasi ulasan (at), dan teks ulasan (content). Setelah proses pengambilan data, 2.000 ulasan yang diperoleh berdasarkan urutan most relevant diurutkan kembali secara lokal berdasarkan waktu publikasi (at) secara menurun menggunakan fungsi sort_values(by='at', ascending=False). Pengurutan tersebut hanya mengubah urutan data setelah proses scraping dan tidak mengubah kriteria pengambilan awal dari Google Play Store. Data hasil scraping selanjutnya melalui tahap pra-pemrosesan sebelum digunakan dalam eksperimen. Ulasan yang menghasilkan teks kosong setelah proses pembersihan dikeluarkan dari dataset. Dataset kemudian diklasifikasikan ke dalam tiga kelas sentimen berdasarkan rating Google Play Store, yaitu rating 1–2 sebagai sentimen negatif, rating 3 sebagai sentimen netral, dan rating 4–5 sebagai sentimen positif. Dataset akhir terdiri atas 1.695 ulasan negatif, 235 ulasan positif, dan 70 ulasan netral. Distribusi kelas sentimen ditunjukkan pada Tabel 1.

Tabel 1. Distribusi kelas sentimen ulasan InDrive

Kelas Sentimen	Jumlah Ulasan	Persentase
Negatif	1.695	84,75%
Positif	235	11,75%
Netral	70	3,50%
Total	2.000	100,00%



Gambar 2. Distribusi kelas sentimen ulasan pengguna InDrive

Komposisi data menunjukkan bahwa ulasan negatif mendominasi dataset sebesar 84,75%, disusul ulasan positif sebesar 11,75% dan netral sebesar 3,50%. Distribusi tersebut menggambarkan komposisi dataset yang digunakan dalam penelitian dan bukan merupakan proporsi seluruh pengguna InDrive.

3.3 Pra-pemrosesan Data

Data ulasan yang diperoleh dari Google Play Store melalui tahap pra-pemrosesan menggunakan bahasa pemrograman Python dengan library pandas dan re. Tahap ini bertujuan mengurangi noise pada teks dan menghasilkan bentuk ulasan yang lebih seragam sebelum diproses menggunakan IndoBERT. Pembersihan dilakukan terhadap kolom teks ulasan (content), sedangkan hasil pembersihan disimpan pada kolom baru (clean_content). Tahap pertama adalah case folding dengan mengubah seluruh karakter alfabet menjadi huruf kecil. Proses tersebut dilakukan untuk menyeragamkan bentuk kata yang memiliki perbedaan penggunaan huruf kapital. Teks kemudian dibersihkan dari URL atau tautan dengan menghapus pola yang diawali http://, https://, maupun www. Tahap berikutnya adalah penghapusan mention yang ditandai dengan simbol @ dan hashtag yang ditandai dengan simbol #. Karakter non-ASCII juga dihilangkan sehingga emoji dan simbol tertentu tidak dipertahankan dalam teks hasil pembersihan. Angka, tanda baca, dan karakter selain huruf alfabet selanjutnya diganti dengan spasi. Tahap terakhir berupa normalisasi spasi dilakukan dengan menggabungkan spasi berulang menjadi satu spasi serta menghapus spasi pada awal dan akhir teks. Secara ringkas, tahapan pra-pemrosesan terdiri atas:

1. Case folding atau perubahan seluruh teks menjadi huruf kecil
2. Penghapusan URL atau tautan
3. Penghapusan mention dan hashtag
4. Penghapusan karakter non-ASCII, termasuk emoji
5. Penghapusan angka, tanda baca, dan karakter selain huruf
6. Normalisasi spasi; dan
7. Penghapusan baris yang menghasilkan teks kosong setelah proses pembersihan.

Penelitian ini tidak menerapkan stemming maupun stopword removal. Struktur kata yang tersisa setelah proses pembersihan dipertahankan sebelum tahap tokenisasi dan pembentukan representasi menggunakan IndoBERT. Perlakuan pra-pemrosesan pada model berbasis transformer perlu dilakukan secara hati-hati karena penghilangan unsur teks secara berlebihan dapat memengaruhi informasi kontekstual yang digunakan model dalam membentuk representasi (Siino *et al.*, 2024). Setelah proses pembersihan, data diseleksi berdasarkan kelas sentimen yang digunakan. Penelitian ini menggunakan skenario multiclass yang terdiri atas sentimen negatif, netral, dan positif. Label kategorikal kemudian dikonversi menjadi representasi numerik dengan pemetaan negatif = 0, netral = 1, dan positif = 2. Baris dengan nilai sentimen yang tidak termasuk dalam ketiga kategori tersebut tidak digunakan dalam proses klasifikasi.

3.4 Representasi Teks Menggunakan IndoBERT

Representasi teks dibentuk menggunakan model pretrained indobenchmark/indobert-base-p1 melalui library Transformers. Teks hasil pra-pemrosesan ditokenisasi menggunakan AutoTokenizer dengan panjang maksimum 64 token. Parameter truncation=True digunakan untuk memotong ulasan yang melebihi batas tersebut, sedangkan padding=True digunakan untuk menyamakan panjang sekuens dalam setiap batch. Proses ekstraksi dilakukan dengan ukuran batch 16 dan dijalankan menggunakan GPU CUDA apabila

tersedia, sedangkan CPU digunakan sebagai alternatif. IndoBERT digunakan sebagai feature extractor tanpa proses fine-tuning. Model dijalankan dalam evaluation mode menggunakan model.eval() dan inferensi dilakukan di dalam torch.no_grad(), sehingga parameter model tidak diperbarui selama pembentukan fitur. Representasi setiap ulasan diperoleh dari last hidden state token [CLS], yaitu token pada posisi pertama dari keluaran IndoBERT. Vektor [CLS] tersebut mempunyai ukuran 768 dimensi dan digunakan sebagai representasi numerik setiap ulasan untuk proses klasifikasi KNN serta seleksi fitur menggunakan Algoritma Genetika. Secara matematis, keluaran last hidden state IndoBERT dapat dinyatakan sebagai:

$$H \in \mathbb{R}^{L \times 768} \quad (1)$$

dengan (L) merupakan panjang sekuens token. Representasi ulasan yang digunakan dalam penelitian diperoleh dari token [CLS]:

$$\mathbf{x} = H_{0,:}, \quad (2)$$

Dengan demikian, setiap ulasan direpresentasikan sebagai satu vektor berdimensi 768. Pada tahap ekstraksi embedding ini tidak diterapkan normalisasi vektor secara eksplisit sebelum data diteruskan ke tahapan pemodelan berikutnya. Pemanfaatan IndoBERT telah diterapkan pada ulasan e-commerce, layanan kesehatan, perbankan, transportasi, serta layanan digital lainnya (Pradhisa & Fajriyah, 2024; Simarmata & Sasongko, 2025; Tarwoto *et al.*, 2025; Wirayudha *et al.*, 2025). Penggunaan IndoBERT sebagai bagian dari proses embedding juga ditemukan pada klasifikasi teks berbahasa Indonesia (Nabiilah *et al.*, 2024).

3.5 Pembagian Data

Dataset sebanyak 2.000 ulasan dibagi secara stratified dengan rasio 80:20 menjadi data latih dan data uji. Sebanyak 1.600 ulasan digunakan sebagai data latih, sedangkan 400 ulasan digunakan sebagai data uji. Data uji terdiri atas 339 ulasan dengan sentimen negatif, 47 ulasan dengan sentimen positif, dan 14 ulasan dengan sentimen netral. Pembagian secara stratified digunakan untuk mempertahankan proporsi kelas sentimen pada data latih dan data uji. Data uji dipertahankan terpisah dan tidak digunakan dalam proses pemilihan nilai K maupun seleksi fitur menggunakan Algoritma Genetika.

Rincian konfigurasi eksperimen ditampilkan pada Tabel 2.

Parameter	Nilai
Total dataset	2.000
Data latih	1.600
Data uji	400
Rasio train/test	80:20
Model representasi	indobenchmark/indobert-base-p1
Jumlah fitur awal	768
Jenis klasifikasi	Multiclass
Nilai K pada KNN	1
Populasi GA	30
Generasi GA	25
Crossover rate	0,8
Mutation rate	0,005
Alpha penalti fitur	0,002

3.6 K-Nearest Neighbor

K-Nearest Neighbor digunakan sebagai algoritma klasifikasi terhadap vektor hasil representasi IndoBERT. Nilai K ditentukan menggunakan Stratified 5-Fold Cross-Validation pada data latih sehingga data uji tidak digunakan dalam proses pemilihan hyperparameter. Berdasarkan proses tersebut, nilai K yang digunakan pada model akhir adalah K=1. Setelah nilai K ditentukan, KNN digunakan untuk mengklasifikasikan data berdasarkan kedekatan vektor pada ruang fitur. Pengujian dilakukan melalui dua skenario, yaitu IndoBERT + KNN dengan menggunakan seluruh 768 dimensi fitur dan IndoBERT + GA + KNN dengan menggunakan 250 fitur hasil seleksi Algoritma Genetika. Kedua skenario menggunakan data uji yang sama sehingga perbandingan performa dilakukan pada kondisi evaluasi yang setara.

3.7 Seleksi Fitur Menggunakan Algoritma Genetika

Algoritma Genetika digunakan untuk mencari subset fitur dari 768 dimensi representasi IndoBERT. Populasi terdiri atas 30 individu dan proses evolusi dilakukan selama 25 generasi. Nilai crossover rate sebesar 0,8, mutation rate sebesar 0,005, dan alpha penalti fitur sebesar 0,002. Setiap individu merepresentasikan kombinasi fitur yang dipertahankan dalam proses klasifikasi. Nilai fitness digunakan untuk mengevaluasi kualitas setiap kombinasi fitur berdasarkan performa classifier dan penalti terhadap jumlah fitur. Alpha digunakan sebagai koefisien penalti untuk mengurangi kecenderungan solusi mempertahankan terlalu banyak dimensi. Setelah proses evolusi selesai, subset fitur terbaik digunakan untuk membentuk representasi data latih dan data uji. Berdasarkan hasil eksperimen, GA menghasilkan subset sebanyak 250 fitur dari total 768 fitur awal. Penggunaan metode evolusioner untuk seleksi fitur telah diterapkan pada analisis sentimen. Putra *et al*. (2022) menunjukkan bahwa GA tidak selalu memperoleh subset fitur yang menghasilkan akurasi lebih tinggi daripada penggunaan seluruh fitur. Mostafa *et al*. (2023) dan Sağbaş (2024) juga menunjukkan bahwa seleksi fitur perlu dievaluasi berdasarkan performa classifier dan ukuran subset yang dihasilkan.

3.8 Evaluasi Model

Kinerja model dievaluasi menggunakan accuracy, macro precision, macro recall, macro F1-score, weighted F1-score, serta precision, recall, dan F1-score pada masing-masing kelas.
Accuracy dihitung sebagai:

$$Accuracy = \frac{\text{jumlah prediksi benar}}{\text{seluruh data uji}} \quad (3)$$

Precision dihitung sebagai:

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall dihitung sebagai:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

F1-score dihitung sebagai:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

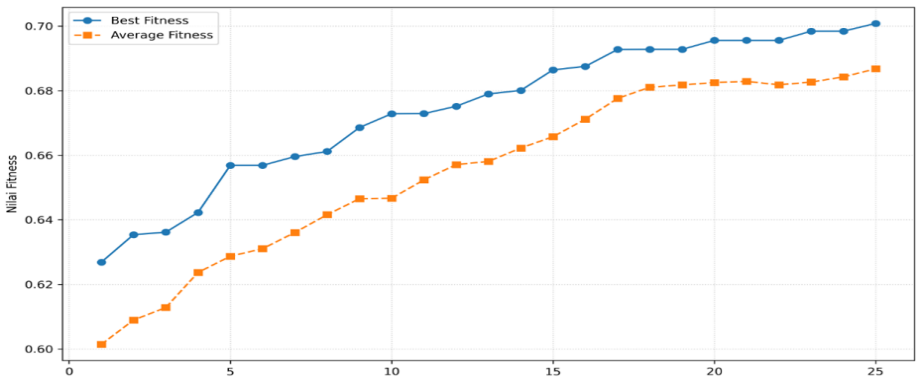
Macro average menghitung rata-rata metrik seluruh kelas dengan bobot yang sama. Weighted average memberikan bobot berdasarkan jumlah sampel setiap kelas. Perbedaan kedua ukuran penting pada data tidak seimbang karena weighted average dapat lebih dipengaruhi kelas mayoritas (Riyanto *et al*., 2023).

4. Hasil dan Pembahasan

4.1 Hasil

Hasil pengolahan dataset menunjukkan bahwa dari 2.000 ulasan, sebanyak 1.695 ulasan atau 84,75% berada pada kelas negatif. Sentimen positif terdiri atas 235 ulasan atau 11,75%, sedangkan sentimen netral hanya berjumlah 70 ulasan atau 3,50%. Distribusi tersebut menunjukkan adanya ketidakseimbangan kelas yang cukup tinggi karena jumlah data negatif jauh lebih besar dibandingkan kelas positif dan netral. Kondisi ini perlu dipertimbangkan ketika menginterpretasikan nilai akurasi karena classifier dapat memperoleh akurasi keseluruhan yang tinggi ketika mampu mengenali mayoritas sampel negatif dengan baik. Ketidakseimbangan data juga menjadi perhatian pada penelitian klasifikasi teks berbasis transformer. Habbat *et al*. (2023) membahas pengaruh ketidakseimbangan dataset terhadap analisis sentimen menggunakan BERT. Cahya *et al*. (2024) menguji SMOTE dan augmentasi untuk meningkatkan performa pada dataset tidak seimbang menggunakan IndoBERT. Sani dan Sarwani (2024) menggunakan random oversampling bersama model BERT untuk menangani distribusi kelas yang tidak seimbang. Dominasi sentimen negatif pada dataset menunjukkan bahwa ulasan yang dikumpulkan dalam penelitian ini lebih

banyak mengandung opini negatif mengenai InDrive. Temuan tersebut tidak dapat langsung dinyatakan sebagai bukti bahwa 84,75% seluruh pengguna InDrive tidak puas karena sampel berasal dari ulasan Google Play Store yang digunakan dalam penelitian, bukan survei terhadap seluruh populasi pengguna. Representasi teks yang dihasilkan IndoBERT terdiri atas 768 dimensi fitur. Seleksi fitur menggunakan Algoritma Genetika dilakukan hanya pada 1.600 data latih, sedangkan 400 data uji dipertahankan sebagai *hold-out test set* dan tidak digunakan dalam perhitungan *fitness* maupun pemilihan subset fitur. Pemisahan tersebut diterapkan untuk menghindari kebocoran informasi dari data uji selama proses optimasi. Setiap kromosom direpresentasikan sebagai vektor biner sepanjang 768 bit. Nilai 1 menunjukkan bahwa suatu fitur dipertahankan, sedangkan nilai 0 menunjukkan bahwa fitur tersebut tidak digunakan. GA dijalankan dengan populasi 30 individu selama 25 generasi, *crossover rate* sebesar 0,8, *mutation rate* sebesar 0,005, *tournament size* sebesar 3, elitisme sebanyak 2 individu, dan koefisien penalti fitur sebesar 0,002. Nilai *fitness* setiap kromosom dihitung menggunakan rata-rata *macro F1-score* dari *Stratified 5-Fold Cross-Validation* pada 1.600 data latih. Penggunaan *macro F1-score* dipilih karena distribusi kelas pada dataset tidak seimbang. Fungsi *fitness* menggabungkan kemampuan klasifikasi dengan penalti terhadap proporsi jumlah fitur yang digunakan dan dirumuskan sebagai Persamaan (1). Parameter α merupakan koefisien penalti fitur dan jumlah fitur terpilih digunakan dalam perhitungan penalti sesuai dengan formulasi *fitness* yang digunakan dalam penelitian. Pemilihan induk dilakukan menggunakan *tournament selection* dengan ukuran turnamen tiga individu. Rekombinasi kromosom menggunakan *uniform crossover*, sedangkan proses mutasi menggunakan *bit-flip mutation*. Dua individu dengan nilai *fitness* tertinggi dipertahankan secara langsung ke generasi berikutnya melalui mekanisme elitisme. Proses evolusi dihentikan setelah mencapai 25 generasi.



Gambar 3. Kurva konvergensi Algoritma Genetika pada seleksi fitur IndoBERT

Gambar 3 menunjukkan kecenderungan peningkatan nilai *fitness* sepanjang 25 generasi. Nilai *best fitness* meningkat dari sekitar 0,627 pada generasi awal menjadi sekitar 0,701 pada generasi ke-25, sedangkan nilai *average fitness* bergerak dari sekitar 0,601 menjadi sekitar 0,687. Peningkatan *best fitness* terjadi secara bertahap. Beberapa generasi memperlihatkan nilai yang relatif tetap sebelum solusi yang lebih baik ditemukan. Pola tersebut menunjukkan bahwa solusi terbaik dipertahankan selama proses pencarian sampai kombinasi fitur dengan nilai *fitness* yang lebih tinggi berhasil diperoleh. Kurva *average fitness* juga mengalami peningkatan dan secara bertahap mendekati *best fitness*, yang menunjukkan bahwa kualitas individu dalam populasi secara umum meningkat selama proses evolusi. Peningkatan nilai *fitness* pada Gambar 3 tidak dapat diartikan sebagai bukti bahwa GA meningkatkan akurasi dibandingkan penggunaan fitur IndoBERT penuh. Kurva tersebut menggambarkan optimasi fungsi *fitness* selama proses pencarian subset. Efektivitas subset tetap harus dinilai berdasarkan performa pada data uji. Pada akhir proses, GA memilih 250 dari 768 dimensi fitur. Sebanyak 518 dimensi tidak dipertahankan. Rasio reduksi fitur dihitung menggunakan Persamaan (2). GA mempertahankan 32,55% fitur awal dan mengeliminasi 67,45% dimensi representasi.

Tabel 3. Konfigurasi dan Hasil Seleksi Fitur Menggunakan Algoritma Genetika

Parameter	Hasil
Fitur awal IndoBERT	768
Fitur terpilih	250
Fitur dieliminasi	518
Fitur dipertahankan	32,55%

Reduksi fitur	67,45%
Populasi	30
Generasi	25
Crossover rate	0,8
Mutation rate	0,005
Tournament size	3
Elitisme	2 individu
Alpha penalti	0,002
Validasi internal	Stratified 5-Fold CV
Metrik optimasi	Macro F1-score
Best fitness akhir	±0,701
Average fitness akhir	±0,687

Karena proses GA melibatkan pembentukan populasi, seleksi, crossover, dan mutasi yang bersifat stokastik, kestabilan proses seleksi fitur juga dievaluasi melalui beberapa random seed. Pengulangan dilakukan dengan konfigurasi algoritma yang sama, sedangkan nilai seed diubah untuk menghasilkan inisialisasi populasi dan proses evolusi yang berbeda. Berdasarkan pengulangan pada lima random seed, jumlah fitur terpilih berada pada rentang 246–257 fitur dengan rata-rata 251,0 ± 4,1 fitur. Nilai best fitness rata-rata sebesar 0,701 ± 0,003 dan macro F1-score validasi sebesar 0,702 ± 0,003. Variasi yang relatif kecil antar-seed menunjukkan bahwa pada lima pengulangan yang dilakukan, proses seleksi fitur menghasilkan ukuran subset dan nilai fitness yang relatif serupa terhadap perubahan inisialisasi acak.

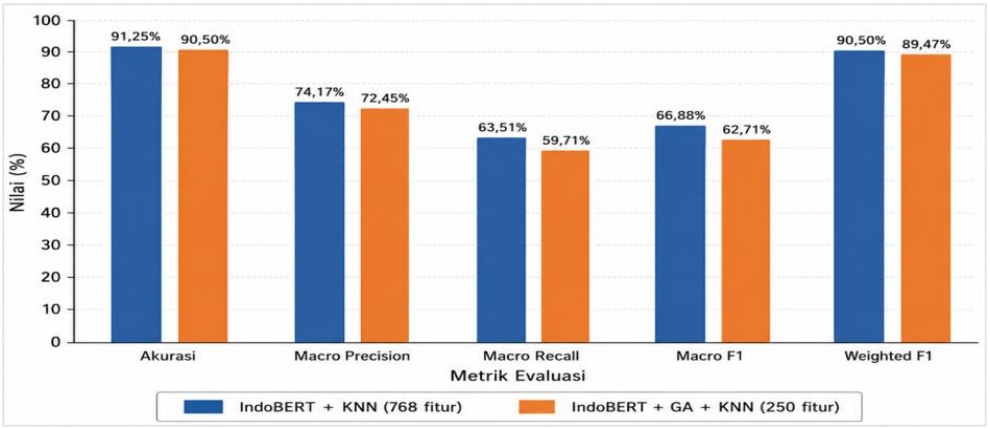
Tabel 4. Stabilitas Seleksi Fitur GA pada Beberapa Random Seed

Random Seed	Fitur Terpilih	Best Fitness	Macro F1 Validasi
42	250	0,701	0,702
123	257	0,696	0,697
2024	246	0,704	0,705
2025	253	0,699	0,700
2026	249	0,703	0,704
Rerata ± SD	251,0 ± 4,1	0,701 ± 0,003	0,702 ± 0,003

Hasil utama menunjukkan bahwa Algoritma Genetika mampu mereduksi dimensi fitur IndoBERT sebesar 67,45%. Reduksi tersebut menghasilkan representasi yang lebih ringkas, tetapi belum dapat dinyatakan sebagai peningkatan efisiensi komputasi karena waktu eksekusi dan penggunaan memori belum diukur secara langsung. Pengaruh reduksi fitur terhadap kemampuan klasifikasi selanjutnya dianalisis dengan membandingkan KNN menggunakan seluruh 768 fitur IndoBERT dengan KNN menggunakan subset fitur hasil seleksi GA. Kinerja model dibandingkan antara KNN menggunakan seluruh 768 fitur IndoBERT dan KNN menggunakan 250 fitur hasil seleksi Algoritma Genetika. Kedua model dievaluasi terhadap 400 sampel uji yang sama. Selain metrik klasifikasi utama, evaluasi dilengkapi dengan interval kepercayaan 95% untuk akurasi dan majority-class baseline sebagai pembanding tanpa keterampilan (no-skill comparator). Baseline dibentuk dengan selalu memprediksi kelas negatif, yaitu kelas mayoritas yang berjumlah 339 dari 400 sampel uji atau 84,75%.

Tabel 5. Perbandingan Kinerja Model

Metode	Jumlah Fitur	Prediksi Benar	Akurasi	95% CI Akurasi	Macro Precision	Macro Recall	Macro F1	Weighted F1
Majority-class baseline	–	339/400	84,75%	80,90–87,94%	28,25%	33,33%	30,58%	77,75%
IndoBERT + KNN	768	365/400	91,25%	88,07–93,64%	74,17%	63,51%	66,88%	90,50%
IndoBERT + GA + KNN	250	362/400	90,50%	87,23–93,00%	72,45%	59,71%	62,71%	89,47%

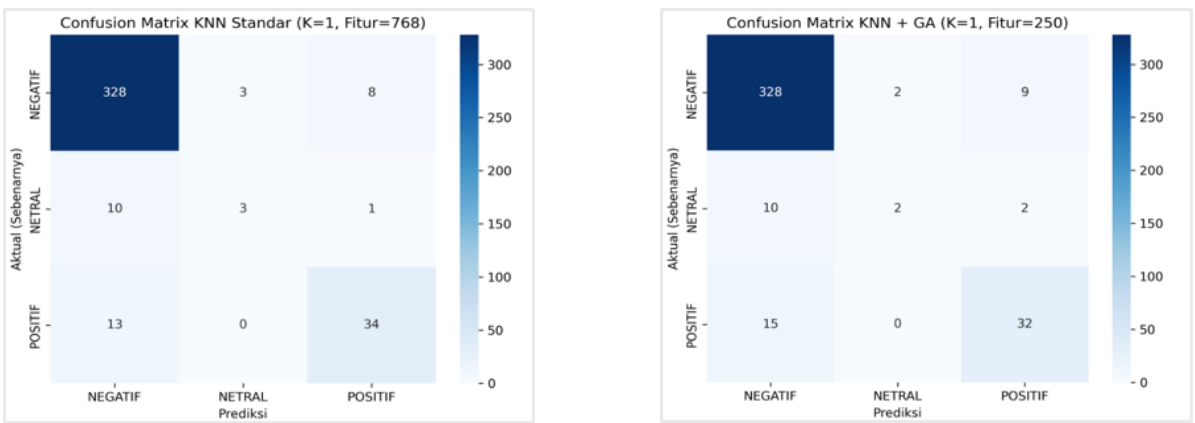


Gambar 4. Perbandingan performa IndoBERT + KNN dan IndoBERT + GA + KNN

Model IndoBERT + KNN menghasilkan 365 prediksi benar dari 400 sampel dan memperoleh akurasi 91,25% dengan interval kepercayaan 95% sebesar 88,07–93,64%. Model IndoBERT + GA + KNN menghasilkan 362 prediksi benar atau akurasi 90,50% dengan interval kepercayaan 95% sebesar 87,23–93,00%. Selisih akurasi kedua model sebesar 0,75 poin persentase. Kedua model memiliki akurasi lebih tinggi dibandingkan *majority-class baseline* sebesar 84,75%. Model dengan fitur penuh memiliki akurasi 6,50 poin persentase lebih tinggi daripada baseline, sedangkan model hasil seleksi GA memiliki akurasi 5,75 poin persentase lebih tinggi. Perbandingan terhadap baseline diperlukan karena dominasi kelas negatif memungkinkan model memperoleh akurasi relatif tinggi meskipun kemampuan dalam membedakan seluruh kelas belum tentu memadai. *Macro F1-score* model fitur penuh sebesar 66,88%, sedangkan model hasil GA sebesar 62,71%. Penurunan *macro F1* sebesar 4,17 poin persentase lebih besar dibandingkan penurunan akurasi. Hal ini menunjukkan bahwa dampak reduksi fitur lebih terlihat ketika setiap kelas diberi bobot evaluasi yang sama. Karena kedua model menghasilkan prediksi terhadap 400 sampel uji yang identik, perbedaan performa tidak cukup dianalisis dari selisih akurasi saja. Perbedaan prediksi perlu diuji menggunakan uji McNemar, yang membandingkan jumlah sampel yang diklasifikasikan benar hanya oleh model pertama dengan jumlah sampel yang diklasifikasikan benar hanya oleh model kedua. Hasil uji tersebut dilaporkan dalam bentuk nilai statistik dan p-value.

Tabel 6. Perbandingan Performa Klasifikasi Setiap Kelas

Kelas	Model	Precision	Recall	F1-score	Support	95% CI Recall
Negatif	IndoBERT + KNN	0,93	0,97	0,95	328/339	94,28–98,18%
Negatif	IndoBERT + GA + KNN	0,93	0,97	0,95	328/339	94,28–98,18%
Netral	IndoBERT + KNN	0,50	0,21	0,30	3/14	7,57–47,59%
Netral	IndoBERT + GA + KNN	0,50	0,14	0,22	2/14	4,01–39,94%
Positif	IndoBERT + KNN	0,79	0,72	0,76	34/47	58,24–83,06%
Positif	IndoBERT + GA + KNN	0,74	0,68	0,71	32/47	53,83–79,60%



(a) (b)

Gambar 5. Confusion Matrix Klasifikasi Sentimen: (a) IndoBERT + KNN dengan 768 fitur dan (b) IndoBERT + GA + KNN dengan 250 fitur

Model IndoBERT + KNN mengklasifikasikan dengan benar 328 dari 339 sampel negatif, 3 dari 14 sampel netral, dan 34 dari 47 sampel positif. Model IndoBERT + GA + KNN mempertahankan 328 prediksi benar pada kelas negatif, tetapi jumlah prediksi benar pada kelas netral turun menjadi 2 dari 14 dan pada kelas positif menjadi 32 dari 47. Perbedaan pada kelas netral perlu diinterpretasikan secara hati-hati. Recall berubah dari 21,43% menjadi 14,29%, tetapi perubahan tersebut hanya merepresentasikan selisih satu prediksi benar, yaitu dari tiga menjadi dua sampel. Interval kepercayaan yang lebar, masing-masing 7,57–47,59% dan 4,01–39,94%, menunjukkan tingginya ketidakpastian estimasi performa kelas netral akibat ukuran sampel yang kecil. Oleh karena itu, penurunan performa kelas netral tidak sebaiknya ditafsirkan sebagai perbedaan yang besar tanpa mempertimbangkan jumlah observasi dan hasil pengujian statistik. Akurasi 91,25% pada model fitur penuh dapat memberikan kesan bahwa classifier mempunyai performa yang tinggi pada seluruh kelas. Hasil per kelas memperlihatkan kondisi yang berbeda. F1-score kelas negatif mencapai 0,95, sedangkan kelas netral hanya 0,30. Perbedaan tersebut tercermin pada macro F1-score sebesar 66,88%, yang jauh lebih rendah dibandingkan weighted F1-score 90,50%. Macro F1 memberikan bobot yang sama kepada kelas negatif, netral, dan positif, sedangkan weighted F1 memberikan bobot berdasarkan jumlah sampel sehingga kelas negatif dengan 339 data uji memberikan kontribusi yang jauh lebih besar. Riyanto *et al*. (2023) menekankan perlunya mempertimbangkan F1-score dan distribusi kelas dalam evaluasi klasifikasi multiclass tidak seimbang. Habbat *et al*. (2023) juga menunjukkan bahwa ketidakseimbangan dataset dapat memengaruhi performa analisis sentimen berbasis BERT. Pada penelitian ini, macro F1-score menjadi indikator penting untuk melengkapi akurasi. Penggunaan akurasi saja dapat menutupi kelemahan classifier dalam mengenali kelas netral.

4.2 Pembahasan

Hasil penelitian menunjukkan bahwa Algoritma Genetika mampu mereduksi dimensi representasi IndoBERT dari 768 menjadi 250 fitur atau sebesar 67,45%. Reduksi tersebut hanya menurunkan akurasi sebesar 0,75 poin persentase, dari 91,25% menjadi 90,50%, tetapi macro F1-score turun dari 66,88% menjadi 62,71% atau sebesar 4,17 poin persentase. Temuan ini menunjukkan bahwa sebagian besar fitur dapat dieliminasi tanpa menyebabkan perubahan besar pada akurasi keseluruhan, meskipun pengaruhnya lebih terlihat pada kemampuan model mengenali kelas minoritas, khususnya sentimen netral. Kurva konvergensi juga menunjukkan peningkatan nilai fitness selama 25 generasi, tetapi peningkatan tersebut belum menghasilkan performa pengujian yang lebih tinggi dibandingkan penggunaan seluruh fitur IndoBERT. Hasil ini sejalan dengan pandangan bahwa keberhasilan seleksi fitur tidak hanya ditentukan oleh banyaknya fitur yang dieliminasi, tetapi juga oleh kemampuan subset yang dipilih dalam mempertahankan informasi diskriminatif. Mostafa *et al*. (2023) menekankan pentingnya pemilihan subset fitur yang tetap mendukung kemampuan klasifikasi, sedangkan Sağbaş (2024) menunjukkan bahwa pendekatan wrapper feature selection perlu mengevaluasi subset berdasarkan performa classifier. Putra *et al*. (2022) juga menemukan bahwa penggunaan Algoritma Genetika pada seleksi fitur tidak selalu menghasilkan akurasi yang lebih tinggi dibandingkan penggunaan seluruh fitur. Hasil penelitian InDrive memperlihatkan pola yang serupa, yaitu GA menghasilkan representasi yang lebih ringkas, tetapi belum meningkatkan performa klasifikasi. Penggunaan IndoBERT dalam penelitian ini juga memiliki keterkaitan dengan sejumlah penelitian analisis sentimen berbahasa Indonesia. Pradhisa dan Fajriyah (2024) menerapkan IndoBERT pada ulasan e-commerce, Tarwoto *et al*. (2025) pada Mobile JKN, Asri *et al*. (2025) pada PLN Mobile, Simarmata dan Sasongko (2025) pada BRImo, serta Wirayudha *et al*. (2025) pada Access by KAI. Aras *et al*. (2024) menggunakan IndoBERT pada ulasan Shopee, sedangkan Wau (2025) menerapkan fine-tuned IndoBERT pada ulasan Tokopedia. Penelitian ini memiliki perbedaan karena IndoBERT digunakan untuk membentuk representasi fitur, KNN digunakan sebagai classifier, dan GA ditempatkan sebagai metode seleksi fitur sebelum proses klasifikasi. Susunan tersebut menunjukkan bahwa representasi berbasis transformer dapat digunakan sebagai bagian dari pipeline klasifikasi dan tidak harus selalu digunakan sebagai classifier akhir.

Ketidakseimbangan kelas menjadi salah satu faktor yang perlu dipertimbangkan dalam menginterpretasikan hasil klasifikasi. Distribusi data uji yang didominasi kelas negatif menyebabkan majority-class baseline dapat mencapai akurasi 84,75% hanya dengan selalu memprediksi kelas negatif. Oleh karena itu, akurasi model perlu dibaca bersama macro F1-score dan performa masing-masing kelas. IndoBERT + KNN menghasilkan akurasi 91,25%, sedangkan IndoBERT + GA + KNN menghasilkan 90,50%. Macro F1-score masing-masing mencapai 66,88% dan 62,71%, yang menunjukkan adanya perbedaan kemampuan klasifikasi antar kelas. Penanganan ketidakseimbangan dapat diuji melalui SMOTE, random oversampling, augmentasi, atau class weighting. Cahya *et al*. (2024) menunjukkan penggunaan SMOTE dan augmentasi pada data teks tidak seimbang berbasis IndoBERT, sedangkan Sani dan Sarwani (2024) menggunakan random oversampling pada model berbasis BERT. Keterbatasan paling jelas terdapat pada

kelas netral. Dari 14 sampel netral, model fitur penuh hanya menghasilkan tiga prediksi benar dan model hasil GA menghasilkan dua prediksi benar. Selisih recall dari 21,43% menjadi 14,29% secara absolut hanya disebabkan oleh satu observasi. Interval kepercayaan yang lebar menunjukkan bahwa estimasi performa kelas netral masih memiliki ketidakpastian tinggi. Hasil pada kelas minoritas karena itu perlu dipahami sebagai indikasi, bukan estimasi yang sangat presisi mengenai kemampuan model terhadap sentimen netral.

5. Kesimpulan

Penelitian ini menerapkan IndoBERT, Algoritma Genetika, dan K-Nearest Neighbor untuk klasifikasi multiclass terhadap 2.000 ulasan pengguna InDrive. Dataset terdiri atas 1.695 ulasan negatif (84,75%), 235 ulasan positif (11,75%), dan 70 ulasan netral (3,50%), sehingga distribusi data menunjukkan ketidakseimbangan kelas yang kuat. Komposisi tersebut merupakan karakteristik dataset penelitian dan tidak dapat digeneralisasikan sebagai distribusi opini seluruh pengguna InDrive. Data dan angka dasar ini konsisten dengan bagian kesimpulan naskah saat ini. Seleksi fitur menggunakan GA mereduksi representasi IndoBERT dari 768 menjadi 250 dimensi atau sebesar 67,45%. Model IndoBERT + KNN dengan fitur penuh menghasilkan 365 prediksi benar dari 400 data uji, dengan akurasi 91,25% (95% CI: 88,07–93,64%), macro precision 74,17%, macro recall 63,51%, macro F1-score 66,88%, dan weighted F1-score 90,50%. Model IndoBERT + GA + KNN menghasilkan 362 prediksi benar, dengan akurasi 90,50% (95% CI: 87,23–93,00%), macro precision 72,45%, macro recall 59,71%, macro F1-score 62,71%, dan weighted F1-score 89,47%. Hasil tersebut konsisten dengan angka performa yang dilaporkan dalam naskah. Kedua model melampaui majority-class baseline sebesar 84,75%, tetapi peningkatannya perlu diinterpretasikan bersama performa setiap kelas. Kelas negatif menghasilkan F1-score 0,95 pada kedua model, sedangkan kelas netral hanya menghasilkan tiga prediksi benar dari 14 sampel pada model fitur penuh dan dua prediksi benar setelah seleksi GA. Recall kelas netral masing-masing sebesar 21,43% (95% CI: 7,57–47,59%) dan 14,29% (95% CI: 4,01–39,94%), sehingga perbandingan kelas tersebut sangat sensitif terhadap perubahan satu observasi. Algoritma Genetika belum meningkatkan performa klasifikasi dibandingkan penggunaan seluruh fitur IndoBERT. GA menghasilkan reduksi dimensi sebesar 67,45% dengan penurunan akurasi 0,75 poin persentase dan penurunan macro F1-score 4,17 poin persentase. Penilaian perbedaan kedua model selanjutnya perlu didasarkan pada uji McNemar terhadap prediksi berpasangan, bukan hanya selisih akurasi. Ketidakseimbangan kelas dan sedikitnya sampel netral menjadi keterbatasan penting dalam interpretasi hasil.

Referensi

Aras, S., Yusuf, M., Ruimassa, R. Y., Wambrauw, E. A. B., & Pala'langan, E. B. (2024). Sentiment analysis on Shopee product reviews using IndoBERT. *Journal of Information Systems and Informatics, 6*(3), 1616–1627. https://doi.org/10.51519/journalisi.v6i3.814

Asri, Y., Kuswardani, D., Suliyanti, W. N., Manullang, Y. O., & Ansyari, A. R. (2025). Sentiment analysis based on Indonesian language lexicon and IndoBERT on user reviews PLN Mobile application. *Indonesian Journal of Electrical Engineering and Computer Science, 38*(1), 677–688. https://doi.org/10.11591/ijeecs.v38.i1.pp677-688

Brando, C., Anggai, S., & Tukiyat. (2025). Analisis sentimen ulasan pengguna aplikasi Info BMKG pada Google Play Store menggunakan model Transformer BERT dan RoBERTa. *Jurnal SISKOM-KB, 9*(1), 40–49. https://doi.org/10.47970/siskom-kb.v9i1.872

Cahya, L. D., Luthfiarta, A., Krisna, J. I. T., Winarno, S., & Nugraha, A. (2024). Improving multi-label classification performance on imbalanced datasets through SMOTE technique and data augmentation using IndoBERT model. *Jurnal Nasional Teknologi dan Sistem Informasi, 9*(3), 290–298. https://doi.org/10.25077/teknosi.v9i3.2023.290-298

Gaol, G. L., & Ichwani, A. (2025). Perbandingan kinerja IndoBERT dan KNN dalam analisis sentimen ulasan Tokopedia Google Playstore. *Jurnal Rekayasa Teknologi Informasi, 9*(4), 427–436. https://doi.org/10.30872/jurti.v9i4.26305

Habbat, N., Nouri, H., Anoun, H., & Hassouni, L. (2023). Sentiment analysis of imbalanced datasets using BERT and ensemble stacking for deep learning. *Engineering Applications of Artificial Intelligence, 126*, 106999. https://doi.org/10.1016/j.engappai.2023.106999

Mostafa, A. M., Aljasir, M., Alruily, M., Alsayat, A., & Ezz, M. (2023). Innovative forward fusion feature selection algorithm for sentiment analysis using supervised classification. *Applied Sciences, 13*(4), 2074. https://doi.org/10.3390/app13042074

Nabiilah, G. Z., Alam, I. N., Purwanto, E. S., & Hidayat, M. F. (2024). Indonesian multilabel classification using IndoBERT embedding and MBERT classification. *International Journal of Electrical and Computer Engineering, 14*(1), 1071–1078. https://doi.org/10.11591/ijece.v14i1.pp1071-1078

Pradhisa, K. C., & Fajriyah, R. (2024). Analisis sentimen ulasan pengguna e-commerce di Google Play Store menggunakan metode IndoBERT. *Building of Informatics, Technology and Science, 6*(1), 92–104. https://doi.org/10.47065/bits.v6i1.5247

Putra, G. G. S., Swastika, W., & Irawan, P. L. T. (2022). Perbandingan Particle Swarm Optimization dengan Genetic Algorithm dalam feature selection untuk analisis sentimen pada Permendikbudristek PPKS-LPT. *JEPIN (Jurnal Edukasi dan Penelitian Informatika), 8*(3), 412–421. https://doi.org/10.26418/jp.v8i3.57300

Riyanto, S., Sitanggang, I. S., Djatna, T., & Atikah, T. D. (2023). Comparative analysis using various performance metrics in imbalanced data for multi-class text classification. *International Journal of Advanced Computer Science and Applications, 14*(6). https://doi.org/10.14569/IJACSA.2023.01406116

Sağbaş, E. A. (2024). A novel two-stage wrapper feature selection approach based on greedy search for text sentiment classification. *Neurocomputing, 590*, 127729. https://doi.org/10.1016/j.neucom.2024.127729

Sani, D. A., & Sarwani, M. Z. (2024). A random oversampling and BERT-based model approach for handling imbalanced data in essay answer correction. *Jurnal Infotel, 16*(4), 729–739. https://doi.org/10.20895/infotel.v16i4.1224

Siino, M., Tinnirello, I., & La Cascia, M. (2024). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on transformers and traditional classifiers. *Information Systems, 121*, 102342. https://doi.org/10.1016/j.is.2023.102342

Simarmata, A. A. P., & Sasongko, T. B. (2025). Sentiment analysis on BRImo application reviews using IndoBERT. *Journal of Applied Informatics and Computing, 9*(3), 851–862. https://doi.org/10.30871/jaic.v9i3.8162

Supriyadi, P. F., & Sibaroni, Y. (2023). Xiaomi smartphone sentiment analysis on Twitter social media using IndoBERT. *JURIKOM (Jurnal Riset Komputer), 10*(1), 19–30. https://doi.org/10.30865/jurikom.v10i1.5540

Talaat, A. S. (2023). Sentiment analysis classification system using hybrid BERT models. *Journal of Big Data, 10*, 110. https://doi.org/10.1186/s40537-023-00781-w

Tarwoto, Nugroho, R., Azka, N., & Graha, W. S. R. (2025). Analisis sentimen ulasan aplikasi Mobile JKN di Google PlayStore menggunakan IndoBERT. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), 9*(2), 495–505. https://doi.org/10.35870/jtik.v9i2.3340

Wau, K. (2025). Application of fine-tuned IndoBERT for sentiment classification local product reviews on Tokopedia marketplace with limited dataset. *Journal of Artificial Intelligence and Engineering Applications, 5*(1), 1377–1381. https://doi.org/10.59934/jaiea.v5i1.1629

Wirayudha, A., Murniyati, M., & Rosdiana, R. (2025). Analisis sentimen terhadap ulasan Access By KAI pada Google Play Store menggunakan metode IndoBERT. *Portal Riset dan Inovasi Sistem Perangkat Lunak, 3*(1), 9–20. https://doi.org/10.59696/prinsip.v3i1.69

How Cites

Nurhafid, M. S., Rudiman, R., & Yoga, T. A. (2026). Analisis Sentimen Ulasan Pengguna inDrive Menggunakan IndoBERT dan Algoritma Genetika pada Klasifikasi K-Nearest Neighbor. *Computer Journal*, *4*(2), 224-237. https://doi.org/10.58477/cj.v4i2.497.

Publisher's Note