Volume 8 Nomor 2 Juli 2026

JJEEE

e-ISSN : 2715-0887
p-ISSN : 2654-7813

Jambura Journal of Electrical and Electronics Engineering

Analisis Komparatif Pendekatan Data-Level dan Cost-Sensitive Learning pada IndoBERT untuk Analisis Sentimen Ulasan Aplikasi Ruangguru

Comparative Analysis of Data-Level and Cost-Sensitive Learning in IndoBERT-Based Sentiment Analysis of Ruangguru App Reviews

Ade Toti Febrian
Teknologi Informasi
Universitas Amikom Purwokerto
Purwokerto, Indonesia
adetotifebrian@gmail.com

Purwadi
Magister Ilmu Komputer
Universitas Amikom Purwokerto
Purwokerto, Indonesia
purwadi@amikompurwokerto.ac.id

Adam Proyogo Kuncoro
Informatika
Universitas Amikom Purwokerto
Purwokerto, Indonesia
adam@amikompurwokerto.ac.id

Diterima : Juni 2026
Disetujui : Juli 2026
Dipublikasi : Juli 2026

***Abstrak*— Ulasan pengguna pada aplikasi pembelajaran daring seperti Ruangguru merupakan sumber informasi penting untuk mengevaluasi kualitas layanan, pengalaman pengguna, serta efektivitas pembelajaran digital. Seiring meningkatnya penggunaan model Transformer, IndoBERT menunjukkan kinerja yang baik dalam analisis sentimen berbahasa Indonesia. Namun, penelitian terdahulu umumnya membandingkan teknik penanganan data tidak seimbang menggunakan dataset, arsitektur model, dan protokol eksperimen yang berbeda sehingga efektivitas relatif antara pendekatan data-level dan cost-sensitive learning pada backbone Transformer yang identik belum dapat dibandingkan secara objektif. Penelitian ini bertujuan melakukan analisis komparatif Pendekatan Data-Level menggunakan Latent-SMOTE dan Pendekatan Cost-Sensitive Learning menggunakan Class-Weighted Loss pada model IndoBERT dalam lingkungan eksperimen yang identik. Dataset terdiri atas 3.767 ulasan aplikasi Ruangguru yang diperoleh melalui Google Play Store dan diproses melalui preprocessing, tokenisasi IndoBERT, pembagian data secara stratified menjadi data latih, validasi, dan pengujian, kemudian dievaluasi menggunakan Accuracy, Precision, Recall, F1-score, confusion matrix, serta uji Cochran's Q dan McNemar untuk menguji signifikansi statistik perbedaan performa model. Hasil penelitian menunjukkan bahwa model Baseline memperoleh Accuracy tertinggi sebesar 90,05%, sedangkan Pendekatan Cost-Sensitive Learning menghasilkan Macro F1-score tertinggi sebesar 0,6275, lebih tinggi dibandingkan Baseline (0,5658) maupun Pendekatan Data-Level (Latent-SMOTE). Temuan tersebut menunjukkan bahwa pembobotan fungsi kerugian mampu meningkatkan kemampuan model dalam mengenali kelas minoritas tanpa mengubah distribusi data pelatihan, sedangkan Latent-SMOTE memberikan peningkatan pada representasi kelas minoritas tetapi belum mampu melampaui performa Class-Weighted Loss. Hasil uji McNemar juga menunjukkan bahwa peningkatan performa terhadap model Baseline signifikan secara statistik. Kontribusi penelitian ini adalah menyajikan evaluasi komparatif menggunakan backbone IndoBERT, dataset, hyperparameter, dan protokol eksperimen yang sama sehingga pengaruh strategi penanganan imbalanced data dapat diamati secara lebih objektif. Selain itu, penelitian ini menerapkan Latent-SMOTE pada representasi laten (latent space) serta melengkapi evaluasi menggunakan pengujian signifikansi statistik, sehingga memberikan dasar empiris yang lebih kuat bagi pengembangan analisis sentimen berbahasa Indonesia.**

Kata Kunci*— *Analisis Sentimen; IndoBERT; Transformer; Natural Language Processing; Ketidakseimbangan Kelas; Imbalanced Learning; Pendekatan Data-Level; Pendekatan Cost-Sensitive Learning; Ulasan Aplikasi Mobile.

Abstract*— *User reviews of online learning applications such as Ruangguru provide valuable information for evaluating service quality, user experience, and digital learning effectiveness. Although IndoBERT has demonstrated strong performance in Indonesian sentiment analysis, previous studies generally compared imbalance handling techniques using different datasets, model architectures, and experimental protocols, making the relative effectiveness of data-level and cost-sensitive learning approaches difficult to evaluate objectively under the same Transformer backbone. This study compares a Data-Level Approach using Latent-SMOTE with a Cost-Sensitive Learning Approach using Class-Weighted Loss on an identical IndoBERT architecture. The dataset consists of 3,767 Ruangguru user reviews collected from Google Play Store and processed through text preprocessing, IndoBERT tokenization, stratified train-validation-test splitting, and evaluation using Accuracy, Precision, Recall, Macro F1-score, confusion matrix, Cochran's Q Test, and McNemar Test. Experimental results show that the Baseline model achieved the highest Accuracy (90.05%), while the Cost-Sensitive Learning approach obtained the highest Macro F1-score (0.6275), outperforming both the Baseline (0.5658) and the Data-Level approach. These findings indicate that class-weighted optimization improves minority-class recognition without

modifying the original training distribution, whereas Latent-SMOTE enhances minority representation but does not outperform Class-Weighted Loss. McNemar testing further confirms that the improvements over the Baseline are statistically significant. The main contribution of this work is an objective comparison of Data-Level and Cost-Sensitive Learning approaches using the same IndoBERT backbone, dataset, preprocessing pipeline, hyperparameters, and evaluation protocol. In addition, the study applies Latent-SMOTE in the latent feature space and complements performance evaluation with statistical significance testing, providing stronger empirical evidence for handling imbalanced Indonesian sentiment datasets

Keywords— Sentiment Analysis; IndoBERT; Transformer; Natural Language Processing; Class Imbalance; Imbalanced Learning; Data-Level Approach; Cost-Sensitive Learning Approach; Mobile Application Reviews.

I. PENDAHULUAN

Ruangguru merupakan salah satu platform pembelajaran daring yang banyak digunakan di Indonesia untuk mendukung proses belajar melalui layanan berbasis digital. Peningkatan jumlah pengguna aplikasi turut menghasilkan ulasan dalam jumlah besar di *Google Play Store* yang mencerminkan pengalaman pengguna terhadap kualitas layanan, kemudahan penggunaan, stabilitas aplikasi, maupun efektivitas proses pembelajaran. Informasi tersebut dapat dimanfaatkan sebagai sumber evaluasi bagi pengembang untuk meningkatkan kualitas aplikasi secara berkelanjutan [1]–[4]. Berbeda dengan survei terstruktur, ulasan pengguna merepresentasikan pengalaman nyata (*real-world user feedback*) sehingga mampu memberikan gambaran yang lebih komprehensif mengenai kepuasan maupun keluhan pengguna terhadap layanan aplikasi.

Analisis sentimen merupakan salah satu bidang Natural Language Processing (NLP) yang banyak memanfaatkan model Transformer, khususnya IndoBERT, karena mampu menghasilkan representasi kontekstual yang lebih baik dibandingkan metode berbasis fitur seperti Naïve Bayes, Support Vector Machine (SVM), dan Random Forest [5]–[8], [14], [15], [24], [25]. Namun, tingginya performa IndoBERT belum secara otomatis mengatasi permasalahan imbalanced data, sehingga dominasi kelas mayoritas masih dapat menurunkan kemampuan model dalam mengenali kelas minoritas.

Meskipun demikian, tingginya kemampuan representasi kontekstual tidak secara otomatis mengatasi permasalahan ketidakseimbangan distribusi kelas (*imbalanced data*) yang umum dijumpai pada ulasan aplikasi, di mana kelas positif mendominasi sehingga model cenderung mengalami penurunan kemampuan dalam mengenali kelas minoritas.

Permasalahan imbalanced data menyebabkan Accuracy kurang merepresentasikan performa model secara menyeluruh, sehingga evaluasi perlu mempertimbangkan Precision, Recall, Macro F1-score, dan confusion matrix [9]–[12], [30]. Penanganannya umumnya dilakukan melalui Pendekatan Data-Level yang menyeimbangkan distribusi data dengan oversampling serta Pendekatan Cost-Sensitive Learning yang memberikan bobot lebih besar pada kelas minoritas melalui loss function tanpa mengubah distribusi data pelatihan [10]–[12], [17]–[21].

Untuk memberikan gambaran perkembangan penelitian secara sistematis, Tabel 1 menyajikan ringkasan penelitian terdahulu yang menjadi dasar penyusunan research gap dan novelty penelitian ini.

TABEL 1. RINGKASAN PENELITIAN TERDAHULU

No.	Peneliti	Objek	Metode	Temuan Utama	Keterbatasan
1	Nur Isma et al. (2026) [13]	Klasifikasi opini	SVM, MLP, Data Augmentation	Data augmentation meningkatkan performa klasifikasi dan model MLP memberikan performa lebih tinggi dibandingkan SVM.	Belum menggunakan model *Transformer* serta belum membahas penanganan *imbalanced learning.*
2	R. Ishak dan A. Bengnga (2026) [28]	Kemiripan judul penelitian berbahasa Indonesia	IndoBERT Embedding + Ditto Whitening	Optimasi embedding IndoBERT meningkatkan kualitas representasi semantik sehingga menghasilkan pengukuran kemiripan yang lebih baik.	Berfokus pada *semantic similarity*, bukan analisis sentimen maupun penanganan data tidak seimbang.
3	Belinda Eka Sarah et al. (2025) [37]	Analisis sentimen kendaraan listrik	Fine-tuning IndoBERT dan IndoBERTweet	IndoBERTweet memberikan performa lebih baik dibandingkan IndoBERT pada dataset penelitian.	Belum mengevaluasi pengaruh ketidakseimbangan data maupun pendekatan *Cost-Sensitive Learning.*
4	Wahyu Widyananda et al. (2025) [38]	Analisis sentimen ulasan *e-commerce* Indonesia	Naïve Bayes, SVM, IndoBERT	IndoBERT mengungguli metode machine learning tradisional pada klasifikasi sentimen.	Berfokus pada perbandingan model dan belum membahas strategi penanganan imbalanced data.
5	Akbar et al. (2024) [39]	Analisis sentimen media sosial X	IndoBERT, Synonym Augmentation, Back Translation	Teknik augmentasi berbasis teks meningkatkan performa IndoBERT, dengan Synonym Augmentation memberikan hasil terbaik.	Tidak membandingkan Pendekatan *Cost-Sensitive Learning* maupun Pendekatan *Data-Level.*
6	Muhammad B. Nugroho et al. (2025) [40]	Analisis sentimen kebijakan publik	IndoBERT dan DeBERTa	Model Transformer memberikan performa tinggi pada klasifikasi sentimen multikelas.	Berfokus pada pemilihan arsitektur Transformer dan belum mengkaji strategi penanganan data tidak seimbang.

Berdasarkan Tabel 1, penelitian terdahulu menunjukkan bahwa sebagian besar studi berfokus pada perbandingan arsitektur model Transformer atau penerapan teknik augmentasi data. Namun, belum terdapat penelitian yang secara langsung membandingkan Pendekatan Data-Level dan Cost-Sensitive Learning menggunakan backbone IndoBERT, dataset, konfigurasi hyperparameter, serta protokol evaluasi yang identik. Dengan demikian, perbedaan performa yang dilaporkan pada penelitian sebelumnya belum sepenuhnya dapat diatribusikan pada strategi penanganan imbalanced data karena masih dipengaruhi oleh perbedaan dataset maupun arsitektur model.

Berdasarkan penelitian terdahulu, perbandingan Pendekatan Data-Level dan Cost-Sensitive Learning masih

sulit dilakukan secara objektif karena menggunakan dataset, arsitektur, dan protokol eksperimen yang berbeda. Selain itu, penerapan SMOTE pada input_ids kurang tepat karena merepresentasikan indeks diskret kosakata. Oleh karena itu, penelitian ini menerapkan Latent-SMOTE pada representasi laten ([CLS] embedding) dan membandingkannya dengan Class-Weighted Loss dalam lingkungan eksperimen yang sama.

Kebaruan (*novelty*) penelitian ini terletak pada tiga aspek utama, yaitu: (1) membandingkan Pendekatan Data-Level (Latent-SMOTE) dan Cost-Sensitive Learning (Class-Weighted Loss) menggunakan *backbone IndoBERT, dataset, preprocessing,* hyperparameter, dan skenario eksperimen yang identik; (2) menerapkan Latent-SMOTE pada representasi laten sehingga menghindari kelemahan metodologis penerapan SMOTE pada *input_ids*; serta (3) melengkapi evaluasi performa menggunakan pengujian signifikansi statistik melalui Cochran's Q Test dan McNemar Test sehingga perbedaan performa model tidak hanya dinilai berdasarkan nilai metrik klasifikasi.

Berdasarkan research gap tersebut, penelitian ini bertujuan melakukan analisis komparatif Pendekatan Data-Level dan Cost-Sensitive Learning untuk menangani ketidakseimbangan data pada model IndoBERT dalam analisis sentimen ulasan aplikasi Ruangguru. Seluruh *eksperimen* dilakukan menggunakan dataset, tahapan *preprocessing*, *tokenisasi*, pembagian data, hyperparameter, dan metrik evaluasi yang sama sehingga perbedaan performa yang diperoleh dapat diinterpretasikan sebagai pengaruh dari strategi penanganan imbalanced data yang diterapkan. Hasil penelitian diharapkan memberikan kontribusi empiris bagi pengembangan metode penanganan data tidak seimbang pada analisis sentimen berbahasa Indonesia sekaligus menjadi referensi bagi penelitian *Natural Language Processing* berikutnya.

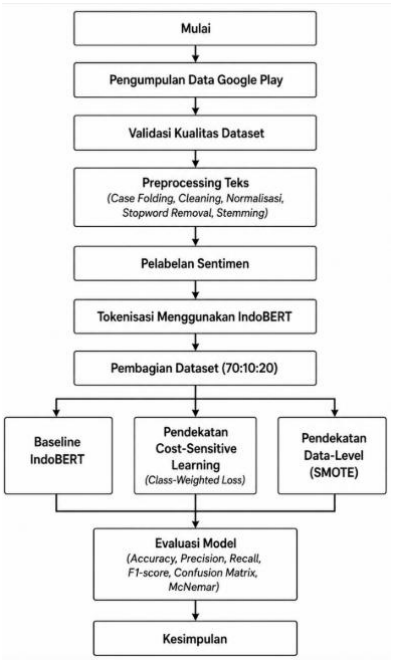
II. METODE

Penelitian ini menggunakan metode eksperimen komparatif untuk mengevaluasi efektivitas dua pendekatan penanganan imbalanced data, yaitu Pendekatan Data-Level menggunakan Latent-SMOTE dan Pendekatan Cost-Sensitive Learning menggunakan Class-Weighted Loss pada model IndoBERT dalam analisis sentimen ulasan aplikasi Ruangguru. Seluruh eksperimen menggunakan dataset, preprocessing, tokenizer, konfigurasi hyperparameter, skenario pelatihan, serta metrik evaluasi yang identik, sehingga perbedaan hasil dapat diatribusikan pada metode penanganan data tidak seimbang yang dibandingkan [11], [12], [26].

Dataset diperoleh melalui web scraping *Google Play Store* menggunakan pustaka *google-play-scraper* terhadap aplikasi Ruangguru pada periode 1 Januari 2025 hingga 29 April 2026 dan menghasilkan 3.767 ulasan. Setelah proses validasi kualitas data, seluruh ulasan diproses melalui tahapan *preprocessing*, pelabelan sentimen, tokenisasi, pembagian data, pelatihan model, dan evaluasi.

A. Alur Penelitian

Untuk memberikan gambaran menyeluruh mengenai tahapan penelitian yang dilakukan, Gambar 1 memperlihatkan alur penelitian mulai dari pengumpulan data hingga evaluasi model. Seluruh tahapan disusun secara sistematis agar proses eksperimen dapat direplikasi pada penelitian selanjutnya.



GAMBAR 1. ALUR PENELITIAN

Berdasarkan Gambar 1, penelitian diawali dengan proses pengumpulan data, validasi dataset, preprocessing, pelabelan sentimen, tokenisasi menggunakan IndoBERT, pembagian dataset secara stratified menjadi data latih, validasi, dan pengujian, kemudian dilanjutkan dengan tiga skenario eksperimen, yaitu Baseline IndoBERT, Pendekatan Cost-Sensitive Learning menggunakan Class-Weighted Loss, dan Pendekatan Data-Level menggunakan Latent-SMOTE. Tahap akhir berupa evaluasi menggunakan metrik klasifikasi dan uji signifikansi statistik untuk memperoleh perbandingan performa yang objektif.

B. Pengumpulan Data

Dataset penelitian berasal dari 3.767 ulasan pengguna Ruangguru yang berhasil dikumpulkan dari *Google Play Store*. Proses validasi menunjukkan bahwa dataset terdiri atas 3.748 pengguna unik, tidak memiliki duplikasi ulasan, tidak terdapat rating maupun tanggal yang tidak valid, sehingga layak digunakan sebagai data penelitian.

Untuk memperlihatkan karakteristik dataset yang digunakan pada penelitian ini, Tabel 2 menyajikan hasil validasi data setelah proses pengumpulan ulasan dari Google Play Store.

TABEL 2. KARAKTERISTIK DATASET HASIL PENGUMPULAN DATA

No.	Karakteristik	Nilai
1	Objek penelitian	Aplikasi Ruangguru (Google Play Store)
2	Periode pengambilan data	1 Januari 2025 – 29 April 2026
3	Total ulasan	3.767
4	Jumlah pengguna unik	3.748
5	Rata-rata rating	4,29
6	Rata-rata panjang ulasan	80,45 karakter
7	Duplikasi ulasan	0
8	Ulasan kosong	0
9	Rating tidak valid	0

No.	Karakteristik	Nilai
10	Tanggal tidak valid	0
11	Status dataset	VALID

Berdasarkan Tabel 2, dataset terdiri atas 3.767 ulasan dari 3.748 pengguna unik dan tidak ditemukan data duplikat maupun data yang tidak valid. Dominasi ulasan dengan rating tinggi mengindikasikan distribusi kelas yang tidak seimbang sehingga diperlukan strategi penanganan imbalanced data sebelum proses pelatihan model.

C. Pelabelan Data

Pelabelan dilakukan menggunakan pendekatan *rule-based labeling* berdasarkan nilai rating pada *Google Play Store*, yaitu rating 1–2 sebagai sentimen negatif, rating 3 sebagai sentimen netral, dan rating 4–5 sebagai sentimen positif. Pendekatan ini dipilih karena rating merepresentasikan tingkat kepuasan pengguna dan telah digunakan secara luas pada penelitian analisis sentimen ulasan aplikasi [1]–[5], [22].

Untuk menjelaskan proses pembentukan label sentimen, Tabel 3 memperlihatkan konversi nilai rating menjadi tiga kategori sentimen yang digunakan pada penelitian.

TABEL 3. SKEMA PELABELAN SENTIMEN

No.	Rating	Label Sentimen	Kode Label
1	1–2	Negatif	0
2	3	Netral	1
3	4–5	Positif	2

Berdasarkan Tabel 3, Skema pelabelan tersebut menghasilkan tiga kelas sentimen, yaitu negatif, netral, dan positif. Distribusi hasil pelabelan menunjukkan dominasi kelas positif sehingga membentuk kondisi data tidak seimbang yang menjadi fokus penelitian ini.

D. Preprocessing Data

Tahap preprocessing meliputi case folding, text cleaning, normalisasi, stopword removal, dan stemming untuk menghasilkan representasi teks yang konsisten sebelum proses tokenisasi.

Untuk memperjelas tahapan *preprocessing* yang diterapkan pada penelitian ini, Tabel 4 menyajikan setiap proses beserta tujuannya.

TABEL 4. TAHAPAN PREPROCESSING DATA

No.	Tahapan	Tujuan
1	Case Folding	Mengubah seluruh huruf menjadi huruf kecil (lowercase).
2	Text Cleaning	Menghapus URL, emoji, tanda baca, angka, karakter khusus, dan spasi berlebih.
3	Normalisasi	Mengubah kata tidak baku menjadi kata baku sesuai kamus normalisasi.
4	Stopword Removal	Menghapus kata umum yang tidak memberikan informasi penting terhadap sentimen.
5	Stemming	Mengubah kata berimbuhan menjadi bentuk kata dasar menggunakan algoritma Sastrawi.

Berdasarkan Tabel 4, preprocessing menghasilkan representasi teks yang lebih bersih dan seragam sehingga variasi penulisan yang tidak relevan dapat diminimalkan sebelum proses tokenisasi. Hasil *preprocessing* menunjukkan tidak terdapat data *duplikat* maupun *missing value*, dengan rata-rata panjang ulasan sebesar 57.9 karakter atau 9.76 *token* per ulasan, sehingga dataset dinyatakan siap digunakan pada tahap *tokenisasi* IndoBERT.

E. Tokenisasi Menggunakan IndoBERT

Tokenisasi dilakukan menggunakan *tokenizer indobenchmark/indobert-base-p1* yang menghasilkan tiga komponen utama, yaitu *input_ids, attention_mask, dan token_type_ids.*

Tokenizer IndoBERT dipilih karena menggunakan *subword tokenization* berbasis *WordPiece* sehingga mampu mengatasi variasi bentuk kata dalam bahasa Indonesia dan mempertahankan informasi semantik pada teks. *Maximum sequence length* ditetapkan sebesar 128 token karena lebih dari 95% ulasan memiliki panjang di bawah batas tersebut berdasarkan hasil analisis distribusi panjang token. Pemilihan nilai tersebut mampu mempertahankan sebagian besar informasi teks sekaligus mengurangi kebutuhan memori GPU dan waktu pelatihan dibandingkan *sequence* yang lebih panjang.

Attention mask digunakan untuk membedakan token yang mengandung informasi dengan *token padding* sehingga mekanisme *self-attention* hanya memproses token yang relevan selama proses *fine-tuning*.

F. Pembagian Dataset

Dataset dibagi secara stratified menjadi 70% data latih, 10% data validasi, dan 20% data pengujian sehingga proporsi masing-masing kelas tetap terjaga pada setiap subset.

Validasi digunakan untuk memilih model terbaik berdasarkan nilai validation loss dan Macro F1-score, sedangkan data pengujian hanya digunakan pada evaluasi akhir agar tidak terjadi kebocoran informasi selama proses pelatihan.

G. Konfigurasi Fine-Tuning IndoBERT

Agar konfigurasi eksperimen dapat direplikasi, Tabel 5 menyajikan hyperparameter yang digunakan pada seluruh skenario pelatihan.

TABEL 5. KONFIGURASI HYPERPARAMETER PELATIHAN MODEL

Parameter	Nilai
Model	indobenchmark/indobert-base-p1
Maximum Sequence Length	128
Batch Size	16
Learning Rate	2×10^{-5}
Epoch	5
Optimizer	AdamW
Scheduler	Linear Scheduler
Random Seed	42

Berdasarkan Tabel 5, seluruh hyperparameter diterapkan secara konsisten pada ketiga skenario eksperimen sehingga setiap perbedaan performa dapat diatribusikan pada

metode penanganan data tidak seimbang, bukan akibat perubahan konfigurasi pelatihanndingkan.

Selain itu, mekanisme *early stopping* diterapkan dengan memantau validation loss pada setiap epoch. Pelatihan dihentikan ketika validation loss tidak lagi membaik, dan model dengan nilai validation loss terendah dipilih sebagai model terbaik untuk mengurangi risiko overfitting serta mempertahankan kemampuan generalisasi.

H. Skenario Eksperimen

Untuk memperjelas konfigurasi eksperimen yang dibandingkan, Tabel 6 menyajikan tiga skenario penelitian.

TABEL 6. SKENARIO EKSPERIMEN

Model	Strategi
Baseline	Fine-tuning IndoBERT tanpa penanganan imbalanced data
Cost-Sensitive Learning	Fine-tuning IndoBERT menggunakan Class-Weighted Loss

Pada Tabel 6 berisi tiga skenario menggunakan *dataset, preprocessing, tokenizer, hyperparamete*r, serta prosedur evaluasi yang sama sehingga setiap perbedaan performa dapat di*interpretasikan* sebagai pengaruh dari strategi penanganan data tidak seimbang yang diterapkan.

F. Evaluasi Model

Evaluasi dilakukan menggunakan *Accuracy, Precision, Recall, Macro F1-score*, dan *confusion matrix*. *Accuracy* tetap digunakan karena memberikan gambaran performa keseluruhan model, sedangkan Macro F1-score menjadi metrik utama karena memberikan bobot yang sama pada setiap kelas sehingga lebih representatif untuk mengevaluasi data tidak seimbang.

Selain metrik klasifikasi, penelitian ini menggunakan *Cochran's Q Test, dan McNemar Test* untuk mengevaluasi apakah perbedaan performa antarmodel signifikan secara statistik sehingga kesimpulan penelitian tidak hanya didasarkan pada selisih nilai metrik.

III. HASIL DAN PEMBAHASAN

A. Distribusi Sentimen Datasset

Untuk memberikan gambaran mengenai distribusi kelas setelah proses pelabelan, Tabel 5 menyajikan jumlah data pada setiap kategori sentimen yang digunakan dalam penelitian. Tabel 6 menyajikan distribusi kelas sentimen hasil proses pelabelan yang digunakan sebagai dataset pada seluruh tahapan eksperimen.

TABEL 7. DISTRIBUSI DATA SENTIMRN

Label Sentimen	Jumlah Data	Persentase
Negatif	555	14,73%
Netral	165	4,38%
Positif	3.047	80,89%
Total	3.767	100,00%

Berdasarkan Tabel 7, distribusi data menunjukkan bahwa kelas Positive mendominasi dataset dengan proporsi sebesar 80,89%, sedangkan kelas *Negative* hanya sebesar 14,73% dan kelas Neutral sebesar 4,38%. Distribusi tersebut mengindikasikan terjadinya ketidakseimbangan kelas (*imbalanced data*) karena sebagian besar data berada pada kelas mayoritas.

Kondisi tersebut berpotensi menyebabkan model pembelajaran mesin lebih banyak mempelajari karakteristik kelas positif sehingga kemampuan mengenali kelas minoritas menjadi menurun. Akibatnya, nilai *Accuracy* dapat terlihat tinggi meskipun model masih mengalami kesalahan klasifikasi pada kelas minoritas. Oleh karena itu, penelitian ini menggunakan Macro F1-score sebagai metrik utama karena memberikan bobot yang sama pada seluruh kelas sehingga lebih representatif dalam mengevaluasi performa model pada dataset yang tidak seimbang.

B. Hasil Pelatihan Model

Untuk membandingkan performa ketiga skenario eksperimen, Tabel 8 menyajikan hasil evaluasi model menggunakan *Accuracy, Precision, Recall, dan Macro F1-score*.

TABEL 8. HASIL EVALUASI MODEL BASELINE INDOBERT

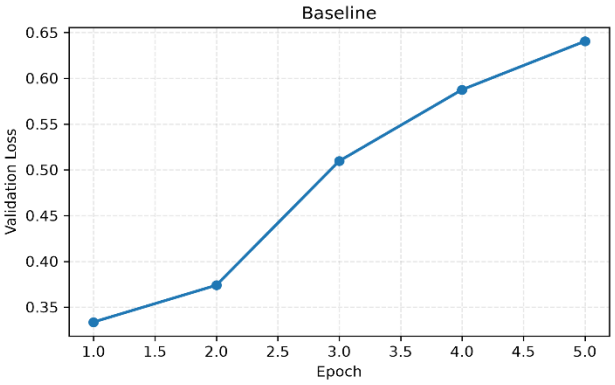
Metric	Baseline	Class-Weighted Loss	Latent-SMOTE	Model Terbaik
Accuracy	0,9005	0,8541	0,8674	Baseline
Precision	0,5465	0,6019	0,6411	Latent-SMOTE
Recall	0,5872	0,6702	0,6822	Latent-SMOTE
Macro F1-score	0,5658	0,6275	0,6397	Latent-SMOTE

Berdasarkan Tabel 8, Pendekatan Cost-Sensitive Learning menghasilkan nilai Macro F1-score tertinggi meskipun bukan Accuracy tertinggi. Temuan tersebut menunjukkan bahwa pembobotan fungsi kerugian lebih efektif meningkatkan kemampuan model dalam mengenali kelas minoritas dibandingkan hanya meningkatkan akurasi keseluruhan.

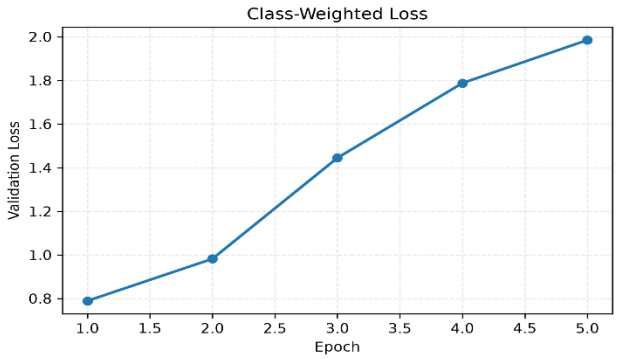
Pendekatan Latent-SMOTE menghasilkan nilai Precision, Recall, dan Macro F1-score tertinggi, menunjukkan bahwa pembangkitan sampel sintetis pada ruang laten mampu memperkaya representasi kelas minoritas. Sementara itu, Class-Weighted Loss juga meningkatkan Recall dan Macro F1-score dibandingkan Baseline melalui pembobotan fungsi kerugian yang mengurangi bias model terhadap kelas mayoritas selama proses optimasi.

C. Analisis Validation Loss

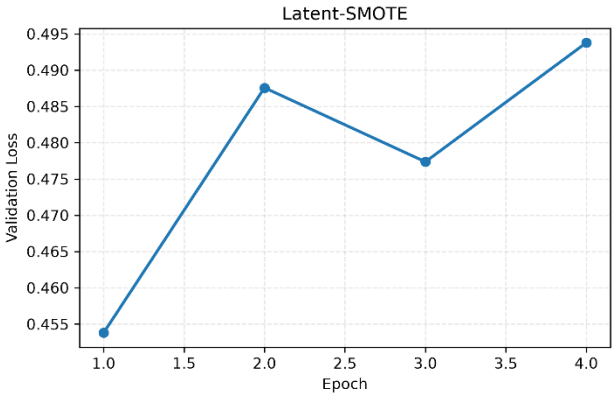
Untuk memperlihatkan proses konvergensi model selama pelatihan, Gambar 2 menyajikan perubahan nilai validation loss pada setiap epoch.



(a) Baseline



(b) Class-Weighted Loss



(c) Latent-SMOTE

GAMBAR 2. PERBANDINGAN KURVA VALIDATION LOSS KETIGA MODEL: (A) BASELINE, (B) CLASS-WEIGHTED LOSS, DAN (C) LATENT-SMOTE

Berdasarkan Gambar 2(a), model Baseline menunjukkan peningkatan validation loss dari sekitar 0,33 menjadi 0,64, yang mengindikasikan menurunnya kemampuan generalisasi dan munculnya kecenderungan overfitting pada epoch akhir. Gambar 2(b) memperlihatkan bahwa Class-Weighted Loss mengalami peningkatan validation loss yang lebih tajam hingga mendekati 2,00, menunjukkan proses optimasi yang lebih sensitif terhadap kesalahan prediksi setelah penerapan pembobotan kelas. Sebaliknya, Gambar 2(c) menunjukkan bahwa Latent-SMOTE memiliki validation loss yang relatif stabil pada kisaran 0,45–0,49, sehingga menunjukkan proses pelatihan yang lebih konsisten. Secara keseluruhan, Gambar 2 menunjukkan bahwa Latent-SMOTE menghasilkan kurva validation loss paling stabil, sedangkan Baseline dan terutama Class-Weighted Loss lebih rentan mengalami overfitting. Oleh karena itu, epoch terbaik dipilih berdasarkan nilai validation loss terendah melalui mekanisme early stopping agar model yang digunakan pada tahap pengujian memiliki kemampuan generalisasi yang optimal.

D. Analisis Precision dan Recall

Untuk memberikan gambaran kemampuan model dalam mengenali masing-masing kelas sentimen, Tabel 8 menyajikan nilai Precision dan Recall dari setiap pendekatan yang digunakan.

TABEL 9. PRECISION DAN RECALL MODEL

Model	Precision	Recall
Baseline	0,5465	0,5872
Class-Weighted Loss	0,6019	0,6702
Latent-SMOTE	0,6411	0,6822

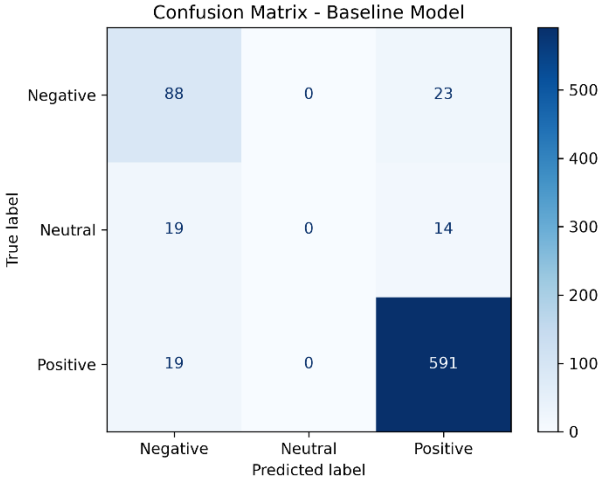
Berdasarkan Tabel 9, model Baseline memiliki nilai Precision dan Recall yang paling rendah dibandingkan kedua pendekatan penanganan data tidak seimbang. Hal tersebut menunjukkan bahwa model masih mengalami kesulitan dalam mengidentifikasi data pada kelas minoritas sehingga menghasilkan lebih banyak kesalahan klasifikasi.

Pendekatan Class-Weighted Loss meningkatkan nilai Recall secara cukup signifikan melalui pemberian bobot yang lebih besar pada kelas minoritas sehingga model lebih mampu mengenali seluruh anggota kelas tersebut.

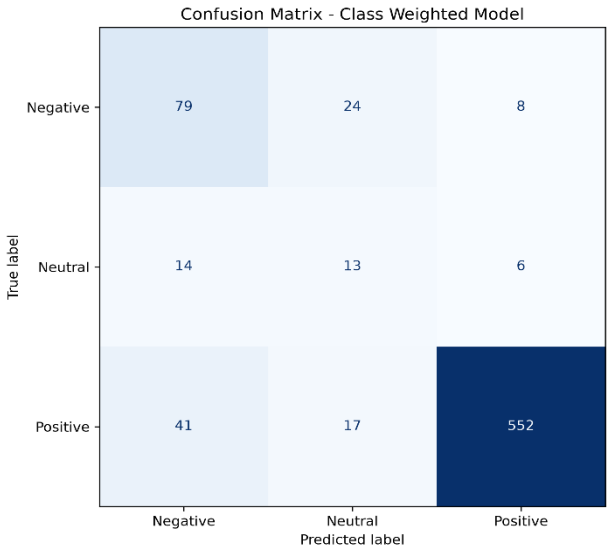
Sementara itu, Latent-SMOTE menghasilkan nilai Precision dan Recall tertinggi. Hasil ini menunjukkan bahwa pembangkitan data sintetis pada ruang laten mampu memperkaya representasi data minoritas sehingga model memiliki kemampuan yang lebih baik dalam membedakan setiap kelas sentimen.

E. Analisis Confusion Matrix Ketiga Model

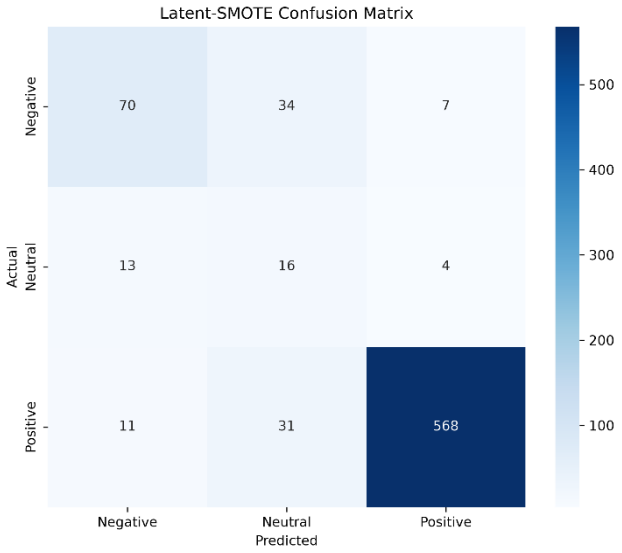
Untuk mengevaluasi pola kesalahan klasifikasi pada setiap pendekatan yang digunakan, hasil prediksi model dibandingkan dengan label sebenarnya menggunakan *confusion matrix*. Perbandingan *confusion matrix* model Baseline, Class-Weighted Loss, dan Latent-SMOTE disajikan pada Gambar 3.



(a) Baseline



(b) Class-Weighted Loss



c) Latent-SMOTE

GAMBAR 3. PERBANDINGAN CUNFUSION MATRIX KETIGA MODEL: (A) BASELINE, (B) CLASS-WEIGHTED LOSS, DAN (C) LATENT-SMOTE

Berdasarkan Gambar 3(a), model Baseline masih didominasi prediksi pada kelas Positive sehingga kemampuan mengenali kelas minoritas, terutama Neutral, masih rendah. Kondisi ini menunjukkan bahwa ketidakseimbangan data menyebabkan model cenderung mempelajari pola kelas mayoritas sehingga kesalahan klasifikasi pada kelas minoritas masih tinggi.

Selanjutnya, Gambar 3(b) menunjukkan bahwa penerapan Class-Weighted Loss meningkatkan pengenalan terhadap kelas Negative dan Neutral melalui pemberian bobot yang lebih besar pada kelas minoritas. Meskipun jumlah prediksi benar pada kelas Positive sedikit menurun, distribusi prediksi menjadi lebih seimbang dibandingkan model Baseline, yang menunjukkan berkurangnya bias terhadap kelas mayoritas.

Berbeda dengan kedua model sebelumnya, Gambar 3(c) memperlihatkan bahwa Latent-SMOTE menghasilkan distribusi prediksi yang paling seimbang pada ketiga kelas. Penambahan sampel sintetis pada ruang representasi laten membantu model mempelajari karakteristik kelas minoritas tanpa menurunkan kemampuan klasifikasi pada kelas mayoritas secara signifikan.

Secara keseluruhan, confusion matrix menunjukkan bahwa model Baseline masih mengalami bias terhadap kelas mayoritas, sedangkan Class-Weighted Loss dan Latent-SMOTE mampu meningkatkan pengenalan kelas minoritas melalui mekanisme yang berbeda, yaitu pembobotan fungsi kerugian dan penyeimbangan representasi data. Temuan ini konsisten dengan peningkatan nilai Precision, Recall, dan Macro F1-score yang diperoleh pada tahap evaluasi model.

F. Uji Signifikansi Statistik

Untuk memastikan bahwa perbedaan performa antar model tidak hanya disebabkan oleh variasi data, dilakukan pengujian statistik menggunakan Cochran's Q Test yang dilanjutkan dengan McNemar Test. Hasil pengujian tersebut disajikan pada Tabel 10.

TABEL 10. HASIL UJI SIGNIFIKANSI STATISTIK

COCHRAN'S Q TEST

Statistik	p-value
19,3069	0,000064

PAIRWISE MCNEMMAR TEST

Perbandingan	Statistik	p-value	Keputusan
Baseline vs Class-Weighted Loss	15,8356	0,000069	Signifikan
Baseline vs Latent-SMOTE	9,7627	0,001781	Signifikan
Class-Weighted Loss vs Latent-SMOTE	1,1571	0,282059	Tidak Signifikan

Berdasarkan Tabel 10, hasil Cochran's Q Test menghasilkan nilai p-value = 0,000064 (< 0,05) sehingga terdapat perbedaan performa yang signifikan di antara ketiga model. Oleh karena itu, pengujian dilanjutkan menggunakan McNemar Test. Hasil McNemar menunjukkan bahwa perbedaan antara Baseline dan Class-Weighted Loss, serta antara Baseline dan Latent-SMOTE, signifikan secara statistik. Sebaliknya, perbedaan antara Class-Weighted Loss dan Latent-SMOTE tidak signifikan (p = 0,282059 > 0,05). Temuan ini menunjukkan bahwa penerapan strategi penanganan ketidakseimbangan kelas memberikan peningkatan performa yang nyata dibandingkan model Baseline, sementara perbedaan efektivitas antara kedua pendekatan penanganan ketidakseimbangan tersebut tidak berbeda secara signifikan.

G. Pembahasan

Hasil penelitian menunjukkan bahwa Pendekatan Cost-Sensitive Learning memberikan performa paling konsisten pada data tidak seimbang karena mekanisme pembobotan fungsi kerugian mengarahkan proses optimasi untuk memberikan perhatian lebih besar terhadap kelas minoritas tanpa mengubah distribusi data asli. Temuan ini sejalan dengan penelitian terdahulu yang melaporkan bahwa pembobotan kelas lebih efektif pada model Transformer

dibandingkan teknik oversampling konvensional ketika jumlah data minoritas relatif terbatas [10]–[12], [17]–[21].

Sebaliknya, Pendekatan Data-Level melalui Latent-SMOTE tetap memberikan peningkatan dibandingkan Baseline, namun efektivitasnya dipengaruhi oleh kualitas representasi laten yang digunakan untuk membentuk sampel sintetis. Walaupun pendekatan ini menghindari kelemahan metodologis penerapan SMOTE pada input_ids, variasi data sintetis yang dihasilkan belum sepenuhnya mampu menggambarkan distribusi semantik data asli sehingga peningkatan performanya masih berada di bawah Class-Weighted Loss.

Implikasi penelitian ini menunjukkan bahwa pada analisis sentimen berbahasa Indonesia menggunakan Transformer, pendekatan Cost-Sensitive Learning dapat menjadi pilihan utama ketika distribusi kelas tidak seimbang dan jumlah data terbatas. Sementara itu, Pendekatan Data-Level berbasis representasi laten tetap memiliki potensi untuk dikembangkan lebih lanjut melalui integrasi dengan teknik augmentasi teks atau embedding yang lebih representatif.

Penelitian ini masih memiliki beberapa keterbatasan. Evaluasi hanya dilakukan pada dataset ulasan aplikasi Ruangguru dan satu backbone Transformer, yaitu IndoBERT, sehingga generalisasi hasil pada domain maupun model lain masih memerlukan pengujian lebih lanjut. Selain itu, Pendekatan Data-Level hanya dievaluasi menggunakan Latent-SMOTE pada representasi laten sehingga efektivitasnya dibandingkan teknik augmentasi teks berbasis Transformer belum dianalisis. Oleh karena itu, hasil penelitian ini perlu dipandang sebagai bukti empiris pada konfigurasi eksperimen yang digunakan.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa Pendekatan Cost-Sensitive Learning menggunakan Class-Weighted Loss memberikan kinerja yang lebih konsisten dibandingkan Pendekatan Data-Level menggunakan Latent-SMOTE dalam menangani ketidakseimbangan data pada model IndoBERT untuk analisis sentimen ulasan aplikasi Ruangguru. Peningkatan tersebut terutama terlihat pada kemampuan model dalam mengenali kelas minoritas yang tercermin melalui peningkatan nilai *Macro F1-score*, sementara Pendekatan Data-Level tetap memberikan perbaikan dibandingkan model Baseline meskipun belum mampu melampaui performa Class-Weighted Loss. Penelitian ini juga menjawab research gap mengenai belum adanya evaluasi yang membandingkan Pendekatan Data-Level dan Cost-Sensitive Learning menggunakan backbone IndoBERT, dataset, tahapan preprocessing, konfigurasi hyperparameter, dan protokol evaluasi yang identik. Dengan menggunakan lingkungan eksperimen yang sama, perbedaan performa dapat diinterpretasikan sebagai pengaruh strategi penanganan data tidak seimbang, bukan akibat perubahan arsitektur model ataupun konfigurasi pelatihan. Selain itu, penerapan Latent-SMOTE pada representasi *laten ([CLS] embedding*) memberikan dasar metodologis yang lebih tepat dibandingkan penerapan SMOTE secara langsung pada *input_ids* karena proses interpolasi dilakukan pada ruang fitur kontinu yang tetap mempertahankan informasi semantik. Kontribusi ilmiah penelitian ini adalah menyediakan evaluasi komparatif yang objektif terhadap dua pendekatan penanganan imbalanced data pada model Transformer berbahasa Indonesia dengan dukungan pengujian signifikansi statistik, sehingga hasil penelitian tidak hanya didasarkan pada perbandingan nilai metrik klasifikasi, tetapi juga pada bukti statistik mengenai perbedaan performa model. Temuan ini memberikan referensi empiris bagi peneliti di bidang Natural Language Processing (NLP) Indonesia dalam memilih strategi penanganan data tidak seimbang serta menjadi masukan bagi pengembang sistem analisis sentimen yang menggunakan model Transformer pada data ulasan pengguna.

Sebagai pengembangan penelitian selanjutnya, Pendekatan Data-Level dapat dikombinasikan dengan teknik augmentasi teks berbasis Transformer, seperti *back-translation, contextual augmentation, atau paraphrase generation*, maupun dengan metode *oversampling* lain yang bekerja pada *representasi embedding*. Selain itu, evaluasi dapat diperluas pada dataset dengan tingkat ketidakseimbangan yang berbeda, domain aplikasi lain, serta model Transformer terbaru agar diperoleh pemahaman yang lebih komprehensif mengenai efektivitas strategi penanganan data tidak seimbang pada analisis sentimen berbahasa Indonesia.

UCAPAN TERIMA KASIH

Penulis mengucapkan puji syukur kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penulis juga menyampaikan ucapan terima kasih yang sebesar-besarnya kepada kedua orang tua tercinta yang senantiasa memberikan doa, dukungan, motivasi, dan semangat selama proses penyusunan penelitian ini; kepada dosen pembimbing skripsi di Universitas Amikom Purwokerto yang telah memberikan ilmu, arahan, dan bimbingan sehingga penelitian ini dapat terselesaikan dengan baik.

REFFERENSI

[1] A. R. Putra and D. E. Ratnawati, "Analisis Sentimen Berbasis Aspek pada Aplikasi Mobile Menggunakan Naïve Bayes Berdasarkan Ulasan Pengguna Google Play Store (Studi Kasus: JConnect Mobile)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 2, pp. 293–300, 2025, doi: 10.25126/jtiik.2025127556.

[2] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjeee.v5i1.16830.

[3] Y. A. Mustofa and I. S. K. Idris, "Ensemble Approach to Sentiment Analysis of Google Play Store App Reviews," *Jambura Journal of Electrical and Electronics Engineering*, vol. 6, no. 2, pp. 181–188, 2024, doi: 10.37905/jjeee.v6i2.25184.

[4] A. Hadiyan, M. Rafli, F. Nugroho, and M. F. Luthfia, "Analisis Sentimen Komentar Pengguna terhadap Aplikasi Prime Video di Google Play Store dengan Pendekatan Machine Learning," *Jurnal Teknologi Informasi dan Komputer*, vol. 6, no. 4, pp. 166–173, 2025.

[5] D. Tribuana, U. Usman, and D. Dayanti, "Penerapan Natural Language Processing untuk Analisis Sentimen terhadap Layanan Publik di Media Sosial Twitter," *Jurnal Teknologi dan Bisnis Cerdas*, vol. 1,

no. 1, pp. 28–37, 2025, doi: 10.64476/jtbc.v1i1.3.

[6] G. Fani and S. Medantoro, "Comparative Analysis of IndoBERT and Classic Machine Learning Models for Sentiment Classification of Education Policy on Social Media X," *Journal of Computer Science and Software*, vol. 10, no. 1, pp. 548–557, 2026.

[7] T. Indriyatmoko, M. Rahardi, H. Utama, A. C. Frobenius, and B. Multilanguage, "Comparative Analysis of BERT and LSTM Models for Sentiment Classification of Mobile Game User Reviews," *Journal of Information Technology*, vol. 10, no. 1, pp. 1093–1100, 2026.

[8] A. N. Wahyuni et al., "Implementasi Model IndoBERT dan mBERT untuk Klasifikasi Teks Berbahasa Indonesia," *Jurnal Teknologi Komputer*, vol. 10, no. 2, pp. 2736–2743, 2026.

[9] J. R. Sari, K. Sadik, A. M. Soleh, and C. Suhaeni, "Evaluasi Model Klasifikasi dalam Deteksi Penipuan Transaksi: Studi Kasus pada Data Tidak Seimbang," *Jurnal Gaussian*, vol. 14, no. 2, pp. 565–576, 2025, doi: 10.14710/j.gauss.14.2.565-576.

[10] A. Karim, B. Bangun, S. Prayetno, and M. Afrendi, "Optimasi Prediksi Harga Sawit Menggunakan Teknik Stacking Algoritma Machine Learning dan Deep Learning dengan SMOTE," *Building Informatics, Technology and Science (BITS)*, vol. 7, no. 1, pp. 638–645, 2025, doi: 10.47065/bits.v7i1.7239..

[11] B. Pes and G. Lai, "Cost-Sensitive Learning Strategies for High-Dimensional and Imbalanced Data: A Comparative Study," *PeerJ Computer Science*, vol. 7, p. e832, 2021, doi: 10.7717/peerj-cs.832.

[12] H. Hairani, T. Widiyaningtyas, and D. D. Prasetya, "Addressing Class Imbalance of Health Data: A Systematic Literature Review on Modified Synthetic Minority Oversampling Technique (SMOTE) Strategies," *International Journal of Informatics and Visualization*, vol. 8, no. 3, pp. 1310–1318, 2024, doi: 10.62527/joiv.8.3.2283.

[13] N. Isma, L. T. Syahyaningsih, D. F. Surianto, and N. Fadilah, "Analysis of Opinion Classification on Marriage Based on Support Vector Machine and Multi-Layer Perceptron," Jambura Journal of Electrical and Electronics Engineering, vol. 8, no. 1, pp. 1–7, 2026, doi: 10.37905/jjeee.v8i1.29616.

[14] A. B. S. Br. Sembiring, R. Robet, and L. Hoki, "Comparison of IndoBERT and SVM Performance in Sentiment Analysis of Digital Education Platforms," *Sinkron*, vol. 10, no. 1, pp. 64–74, 2026, doi: 10.33395/sinkron.v10i1.15472.

[15] R. Anadra, H. Wijayanto, and K. Sadik, "Sentiment Analysis of Tokopedia Customer Reviews Using BiLSTM and IndoBERT with Comparative Analysis of Preprocessing and Labeling Methods," *International Journal of Advances in Data and Information Systems*, vol. 6, no. 3, pp. 773–788, 2025, doi: 10.59395/ijadis.v6i3.1458.

[16] R. J. Siger, M. Naufal, and F. Alzami, "Analisis Performa Class Weight dan Focal Loss pada Model IndoBERT untuk Klasifikasi Teks Depresi Berbahasa Indonesia," *JURIKOM (Jurnal Riset Komputer)*, vol. 13, no. 2, pp. 558–568, 2026, doi: 10.30865/jurikom.v13i2.9620.

[17] L. D. Cahya, A. Luthfiarta, J. I. T. Krisna, S. Winarno, and A. Nugraha, "Improving Multi-label Classification Performance on Imbalanced Datasets Through SMOTE Technique and Data Augmentation Using IndoBERT Model," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 3, pp. 290–298, 2024, doi: 10.25077/teknosi.v9i3.2023.290-298.

[18] F. Y. A'la, "Optimasi Klasifikasi Sentimen Ulasan Game Berbahasa Indonesia: IndoBERT dan SMOTE untuk Menangani Ketidakseimbangan Kelas," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 1, pp. 256–265, 2025, doi: 10.29408/edumatic.v9i1.29666.

[19] I. Saputra and G. L. Ginting, "Sentiment Analysis of Tokopedia Customer Reviews Using IndoBERT and SMOTE for Class Imbalance Handling," *Journal of Online Information Systems and Computing*, vol. 7, no. 1, pp. 20–27, 2025, doi: 10.47065/josyc.v7i1.8748.

[20] M. R. Roihan, M. I. Mazdadi, I. Budiman, and F. Indriani, "Comparative Performance of IndoBERT-Based Deep Learning Models with SMOTE for Trade Tariff Sentiment Analysis," *Journal of Big Data Research*, pp. 432–446, 2026..

[21] R. Widayanti and F. C. Kasih, "Comparative Analysis of Baseline IndoBERT, Class-Weighted IndoBERT, and SMOTE with Support Vector Machine for Handling Imbalanced Sentiment Classification in Indonesian," *Jurnal Teknik Informatika (JUTIF)*, vol. 7, no. 3, 2026.

[22] A. S. Agil Rafsanjani, D. L. Fithri, and S. Supriyono, "Sentiment Analysis of User Reviews of the KitaLulus Application on Google Play Store Using the Support Vector Machine (SVM) Algorithm," *Sistemasi*, vol. 14, no. 5, 2025, doi: 10.32520/stmsi.v14i5.5519.

[23] B. Siswanto and M. Hanafi, "Optimasi IndoBERT untuk Pengenalan Entitas Bernama Bahasa Indonesia pada Data Media Sosial dengan Penalaan Hiperparameter Optuna," *JURIKOM (Jurnal Riset Komputer)*, vol. 13, no. 1, pp. 221–233, 2026, doi: 10.30865/jurikom.v13i1.9545.

[24] R. I. Perwira, V. A. Permadi, D. I. Purnamasari, and R. P. Agusdin, "Domain-Specific Fine-Tuning of IndoBERT for Aspect-Based Sentiment Analysis in Indonesian Travel User-Generated Content," *Journal of Information Systems Engineering and Business Intelligence*, vol. 11, no. 1, pp. 30–40, 2025, doi: 10.20473/jisebi.11.1.30-40.

[25] I. Mursidah et al., "Klasifikasi Sentimen Google Play Store Aplikasi ChatGPT Berbahasa Indonesia Berbasis IndoBERT," *Jurnal Minfo Polgan*, vol. 14, no. 2, pp. 3349–3359, 2025, doi: 10.33395/jmp.v14i2.15751.

[26] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *in Proc.*

2019 Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, 2019, pp. 4171–4186.

[27] A. R. Illawati, A. I. Hadiana, and M. Melina, "Klasifikasi Penyakit Monkeypox dengan XGBoost dan SMOTE untuk Penanganan Data Tidak Seimbang," *Jurnal Algoritma*, vol. 22, no. 2, pp. 35–44, 2025, doi: 10.33364/algoritma/v.22-2.2349.

[28] R. Ishak and A. Bengnga, "Optimizing IndoBERT Embedding with Ditto Whitening for Measuring Research Title Similarity," *Jambura Journal of Electrical and Electronics Engineering*, vol. 8, no. 1, pp. 46–54, 2026, doi: 10.37905/jjeee.v8i1.35554.

[29] W. Clarisha, A. A. M. Fani, D. F. Surianto, and N. Fadilah, "Sentiment Analysis of Local Sunscreen Skintific, Somethinc, and Avoskin with Naïve Bayes and SVM," *Jambura Journal of Electrical and Electronics Engineering*, vol. 7, no. 2, pp. 264–271, 2025, doi: 10.37905/jjeee.v7i2.30257.

[30] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, "Accuracy, Precision, Recall, F1-Score, or MCC? Empirical Evidence from Advanced Statistics, Machine Learning, and Explainable AI for Evaluating Business Predictive Models," *Journal of Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01313-4..

[31] A. C. Ghazy, G. Ghozali, and K. A. Wibowo, "Transformasi Pendidikan: Pengembangan Metodologi dan Media Pembelajaran di Era Digital," *Action Research Journal Indonesia*, vol. 7, no. 4, 2025, doi: 10.61227/arji.v7i4.594.

[32] M. A. Aljauhari and F. Arifin, "Implementasi SMOTE pada Klasifikasi Email Spam Menggunakan Algoritma Machine Learning Berbasis TF-IDF," *Jurnal Algoritma*, vol. 23, no. 1, 2026, doi: 10.33364/algoritma.v23i1.3246.

[33] A. N. Laily, M. A. Barata, and D. Nurdiansyah, "Analisis Klasifikasi Hepatitis Menggunakan Synthetic Minority Oversampling Technique, Support Vector Machine, dan Random Forest," *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 6, no. 1, pp. 223–236, 2026, doi: 10.51454/decode.v6i1.1630.

[34] A. N. Putro and M. A. Muslim, "Penanganan Ketidakseimbangan Data Ekstrim pada Sistem Prediksi," *Techno.Com*, vol. 24, no. 4, pp. 1359–1367, 2025, doi: 10.62411/tc.v24i4.15005.

[35] M. A. A. Saputra et al., "IndoBERT-Relevancy: A Context-Conditioned Relevancy Classifier for Indonesian Text," *arXiv preprint* arXiv:2603.26095, pp. 1–9, 2026.

[36] A. Z. Muhabbab, B. Bunyamin, and H. Hasmawati, "Sentiment Analysis of Indonesia's Free Nutritious Meal Program on Platform X (Formerly Twitter) Using IndoBERT," *ZERO: Jurnal Sains, Matematika, dan Terapan*, vol. 9, no. 3, 2025, doi: 10.30829/zero.v9i3.27629.

[37] B. E. S. Dewi et al., "Sentimen Kendaraan Listrik Menggunakan Fine-Tuning IndoBERT dan IndoBERTweet," *Antivirus Journal*, vol. 19, no. 1, pp. 45–56, 2025.

[38] W. Widyananda et al., "Machine Learning and Transformer-Based Model for Indonesian E-Commerce Review Sentiment Analysis," *International Journal of Computer Science*, vol. 14, no. 2, pp. 112–125, 2025.

[39] Akbar et al., "Klasifikasi Sentimen untuk Mengetahui Kecenderungan Opini Politik pada Media Sosial X Berbasis IndoBERT," *Building Informatics, Technology and Science (BITS),* vol. 6, no. 3, pp. 1420–1429, 2024.

[40] M. B. Nugroho et al., "Multiclass Sentiment Analysis of Electric Vehicle Incentive Policy Using IndoBERT and DeBERTa," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 1, pp. 88–97, 2025.