

## Optimasi Klasifikasi Sentimen Ulasan Game Berbahasa Indonesia: IndoBERT dan SMOTE untuk Menangani Ketidakseimbangan Kelas

Fiddin Yusfida A'la<sup>1,\*</sup>

<sup>1</sup> Program Studi Teknik Informatika, Universitas Sebelas Maret, Indonesia

\* Correspondence: fiddin@staff.uns.ac.id

**Copyright:** © 2025 by the authors

Received: 28 Februari 2025 | Revised: 10 Maret 2025 | Accepted: 30 Maret 2025 | Published: 17 April 2025

### Abstrak

Peningkatan penggunaan aplikasi permainan di *platform* seperti *Google Play Store* telah menandai pentingnya ulasan pengguna sebagai sumber evaluasi kualitas aplikasi. Namun, analisis sentimen ulasan berbahasa Indonesia menghadapi tantangan karena kekhasan struktur bahasa, ekspresi emosional, serta penggunaan slang dan istilah khusus dalam ulasan *game*. Penelitian ini bertujuan untuk mengklasifikasikan ulasan ke dalam tiga kelas sentimen: positif, negatif, dan netral menggunakan model *IndoBERT-base-uncased*. Jenis penelitian yang digunakan adalah eksperimen dengan membandingkan kinerja model menggunakan *dataset* asli dan sintetis. Total *dataset* asli yang dikumpulkan sebanyak 998 ulasan. Parameter *k\_neighbors* SMOTE yang digunakan sebesar 5. Parameter *epoch* *IndoBERT-base-uncased* adalah 10 dengan nilai *batch* per perangkat dan *batch* untuk evaluasi adalah 16. Konfigurasi variabel *warmup\_steps* bernilai 500 dengan Regularisasi L2 *weight\_decay* bernilai 0,01. Hasil evaluasi setelah implementasi SMOTE, skor *precision* meningkat dari 0,44 ke 0,45, *F1-score* dari 0,46 menjadi 0,47. Namun, skor *recall* tidak mengalami peningkatan. Hasil evaluasi menunjukkan bahwa model memiliki performa bervariasi antar kelas dengan akurasi awal sebesar 69,5%. Ketidakseimbangan data menjadi tantangan utama, terutama pada kelas minoritas seperti kelas 1 (netral), yang tidak dapat diprediksi oleh model. Teknik SMOTE berhasil meningkatkan keseimbangan data dan meningkatkan akurasi menjadi 72,5%, serta memperbaiki metrik seperti *precision*, *recall*, dan *F1-score* secara keseluruhan.

**Kata kunci:** analisis sentimen; *indobert*; ulasan pengguna; smote; ketidakseimbangan data

### Abstract

The increased use of gaming apps on platforms like the Google Play Store has signaled the importance of user reviews as a source of app quality evaluation. However, sentiment analysis of Indonesian-language reviews faces challenges due to the peculiarities of language structure, emotional expressions, and the use of slang and specialized terms in game reviews. This study aims to classify reviews into three sentiment classes: positive, negative, and neutral, using the *IndoBERT-base-uncased* model. The type of research used is experimental by comparing the performance of the model using original and synthetic datasets. The total original dataset collected was 998 reviews. The *k\_neighbors* SMOTE parameter used is 5. The *IndoBERT-base-uncased* epoch parameter is 10, with a batch value per device and a batch for evaluation of 16. Configuration variable *warmup\_steps* is 500 with L2 *weight\_decay* regularization at 0.01. Evaluation results after SMOTE implementation: the precision score increased from 0.44 to 0.45, and the *F1-score* from 0.46 to 0.47. However, the recall score did not increase. The evaluation results show that the model has variable performance between classes with an initial accuracy of 69.5%. Data imbalance is a major challenge, especially in minority classes such as class 1 (neutral), which cannot be predicted by the model. The SMOTE technique successfully improved data balance and increased accuracy to 72.5%, as well as improving metrics such as *precision*, *recall*, and *F1-score* overall.



**Keywords:** *sentiment analysis; indobert; user reviews; smote; imbalanced dataset*

## PENDAHULUAN

Aplikasi permainan di *platform* seperti *Google Play* telah menjadi salah satu sumber hiburan utama bagi masyarakat. Dengan semakin banyaknya pengguna aplikasi game, platform ini juga menyediakan fitur bagi pengguna untuk memberikan ulasan dan ulasan mengenai pengalaman mereka. Ulasan ini mencerminkan opini dan perasaan pengguna terkait dengan aplikasi game yang mereka mainkan. Oleh karena itu, analisis terhadap sentimen dalam ulasan pengguna sangat penting untuk menilai kualitas aplikasi dan mendapatkan wawasan mengenai kepuasan serta masalah yang dihadapi oleh pengguna.

Salah satu tantangan utama dalam analisis sentimen ulasan *game* adalah bahasa yang digunakan oleh para pemain. Bahasa Indonesia, sebagai bahasa yang dominan digunakan oleh pengguna di Indonesia, memiliki kekhasan tersendiri dalam struktur kalimat dan ekspresi emosional (Cahyono & Sawitri, 2024; Nuradita, 2022). Dalam konteks ulasan *game*, keunikan ini semakin kompleks karena sering kali pengguna menggunakan campuran bahasa formal, bahasa informal, slang, hingga istilah-istilah spesifik terkait game, seperti "*level grinding*," "*lag*," atau "*gameplay*." Selain itu, ekspresi emosi dalam ulasan *game* cenderung lebih bervariasi, mulai dari antusiasme berlebihan hingga kritik pedas, yang memerlukan pemahaman mendalam tentang konteks budaya dan preferensi pengguna lokal. Masalah ini menjadi semakin menantang ketika menggunakan model yang belum dilatih secara khusus untuk menangani ciri-ciri bahasa Indonesia dengan baik.

Di masa lalu, pendekatan *manual* dan *machine learning* menjadi solusi utama untuk mengatasi tantangan tersebut. Salah satu teknik yang sering digunakan adalah *word embedding*, yang mampu merepresentasikan kata-kata dalam ruang vektor semantik. Namun, teknik ini masih memiliki keterbatasan dalam menangkap konteks spesifik ulasan *game*, seperti interpretasi istilah teknis atau ekspresi emosional yang unik dalam komunitas gamer. Oleh karena itu, diperlukan pendekatan yang lebih sensitif terhadap karakteristik bahasa serta domain ulasan game untuk meningkatkan akurasi analisis sentimen. Analisis sentimen bisa diaplikasikan untuk berbagai kasus lain, contohnya untuk menganalisis sentimen pada *review* aplikasi pembelian bahan bakar (A'la, 2022) dan aplikasi kesehatan (Imaduddin, et al., 2023).

Pendekatan yang dilakukan untuk mengatasi permasalahan ulasan berbahasa Indonesia yaitu model *IndoBERT*. Model bahasa ini berbasis *Bidirectional Encoder Representations from Transformers* (BERT). *IndoBERT* adalah model pra-latih berbasis transformer yang telah terbukti efektif dalam berbagai tugas pemrosesan bahasa alami seperti deteksi disinformasi, analisis sentimen, dan identifikasi *hate speech* (Imaduddin, A'la, et al., 2023; Imaduddin, Kusumaningtias, et al., 2023; Pamungkas et al., 2024; Pradnyana et al., 2023; Putra et al., 2023; Rahmawati et al., 2022; Setiawan & Suryono, 2024; Yulfa et al., 2023). *IndoBERT* mempunyai kinerja yang baik karena kemampuannya dalam memahami konteks kata dalam kalimat secara lebih mendalam.

Model *IndoBERT* yang digunakan pada kasus ini yaitu *IndoBERT-base-uncased*. *Indobert-base-uncased* adalah model bahasa yang telah dilatih untuk memahami teks dalam bahasa Indonesia. Model ini merupakan bagian dari keluarga model BERT yang dirancang khusus untuk pemrosesan *Natural Language Processing* (NLP) bahasa Indonesia (Koto et al., 2020; Wilie et al., 2020). *Indobert-base-uncased* merupakan model yang tidak sensitif terhadap huruf besar atau kecil.

Permasalahan utama dalam penelitian ini adalah ketidakseimbangan kelas pada dataset ulasan aplikasi *game*, di mana jumlah ulasan negatif (kelas minoritas) lebih sedikit dibandingkan ulasan positif atau netral. Ketidakseimbangan ini menyebabkan model kesulitan memprediksi ulasan negatif dengan akurat, sehingga persepsi pengguna, seperti keluhan terkait

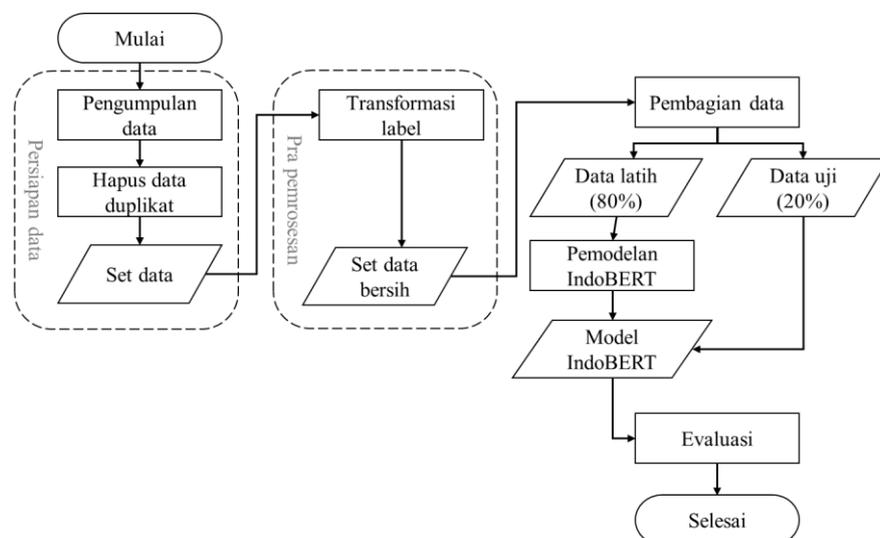
bug atau pengalaman buruk sering terabaikan. Untuk mengatasi masalah ini, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) akan digunakan.

SMOTE bekerja dengan menghasilkan contoh sintetis untuk kelas minoritas berdasarkan data yang ada, sehingga distribusi kelas menjadi lebih seimbang (Chawla et al., 2002). Metode ini telah terbukti efektif dalam meningkatkan performa model pada dataset dengan ketidakseimbangan kelas (A'la et al., 2023; Dharmendra et al., 2024; Hidayat et al., 2021; Irfannandhy et al., 2024; Jimoh et al., 2019; Lv et al., 2018; Mohammed, 2020; Pardede & Pamungkas, 2024). Sebagian besar penelitian sebelumnya menerapkan SMOTE pada data numerik, seperti dataset medis, finansial, atau sensor IoT. Data numerik memiliki struktur yang lebih homogen dan pola hubungan antarfitur yang lebih mudah dipetakan secara matematis. Penelitian-penelitian sebelumnya jarang membahas bagaimana SMOTE bekerja pada data teks, yang memiliki dimensi tinggi dan pola linguistik kompleks. Data teks sering kali mengandung variasi bahasa, seperti sarkasme, dialek, atau frasa ambigu, yang sulit direplikasi oleh SMOTE. Sementara pada penelitian ini, kami akan melakukan eksplorasi SMOTE dan IndoBERT pada data teks.

Penelitian ini bertujuan mengembangkan model klasifikasi sentimen menggunakan *IndoBERT-base-uncased* untuk menganalisis ulasan aplikasi *game* berbahasa Indonesia di *Google Play* menjadi tiga kategori: positif, negatif, dan netral. Fokus utama adalah menangani masalah ketidakseimbangan kelas dengan menerapkan teknik SMOTE, sehingga model lebih sensitif terhadap ulasan negatif dan netral sebagai kelas minoritas. Dengan kombinasi IndoBERT dan SMOTE, diharapkan model dapat mencapai akurasi tinggi dan memberikan wawasan komprehensif bagi pengembang *game* untuk meningkatkan kualitas aplikasi berdasarkan umpan balik pengguna.

## METODE

Desain penelitian ini merupakan eksperimen yang mengevaluasi teknik SMOTE untuk mengatasi bias kelas. Eksperimen dilakukan dengan dua skenario pengujian, yaitu menggunakan data asli dan data sintetis, dengan parameter  $k\_neighbors$  SMOTE sebesar 5. Alur eksperimen terbagi menjadi tiga tahap: persiapan, pra-pemrosesan, dan pemodelan (lihat gambar 1). Tahap persiapan, melibatkan penggunaan media *Google Colaboration Released version 1.2.0* untuk eksperimen, termasuk pengumpulan dan penghapusan data duplikat. Pengumpulan data dilakukan menggunakan library *Google-Play-Scraper version:1.2.7* di platform *Google Play* dengan kriteria berdasarkan username dan ulasan yang identik. Setelah pemeriksaan, tidak ditemukan data duplikat.



**Gambar 1.** Diagram alir eksperimen

Tahap pra-pemrosesan dilakukan transformasi skor ulasan diubah dari skala 1–5 menjadi tiga kategori: 0 (netral), 1 (negatif), dan 2 (positif), kemudian dibagi menjadi data latih (80%) dan data uji (20%). Model *Artificial Intelligence* dilatih menggunakan IndoBERT, dan hasilnya dievaluasi. Penelitian ini menggunakan 998 ulasan berbahasa Indonesia dari permainan *Free Fire*, yang saat itu merupakan *game* paling banyak diunduh. Distribusi ulasan tidak seimbang: 578 ulasan (58%) berbintang 5, 253 ulasan (25%) berbintang 1, sementara bintang 2, 3, dan 4 masing-masing hanya 54, 54, dan 59 ulasan. Mayoritas ulasan terkonsentrasi pada bintang 5 dan 1, sedangkan kelas lainnya jarang muncul. Tabel 1 merupakan contoh ulasan negatif (skor 1–2) seperti “Ga bagus” dan “Jelek”; ulasan positif (skor 4–5) seperti “Gem nya keren bagus” dan “Good”; sedangkan skor 3 bersifat netral hingga sedikit negatif, seperti berisi keluhan.

**Tabel 1.** Contoh ulasan

Skor (Bintang)	Ulasan
1	Ga bagus Jelek
2	FF memberi tim gak bisa main terus tim gue masa nub mulu gak jelas FF ni Gamenya banyak ngebugkk.....!!!
3	Tolong ini yang player cheatnya dibann dong, dari pada rank ku turun trs!!!! Mantul
4	Gem nya kren bagus Good
5	Sangat bagus Bagus

Transformasi skor untuk mengurangi kompleksitas model dengan mengelompokkan 5 kelas menjadi 3 menggunakan Persamaan (1). Hal ini memudahkan model menangkap pola dominan. Label 0, 1, 2 merepresentasikan skala ordinal (negatif < netral < positif), serta memungkinkan model mempelajari hierarki sentimen. Jumlah ulasan hasil penerapan Persamaan (1) yaitu negatif sejumlah 307, netral sejumlah 54, dan positif sejumlah 637. Nilai *epoch* yang digunakan pada *indobert-base-uncased* adalah 10. Nilai *batch* per perangkat dan *batch* untuk evaluasi adalah 16. Konfigurasi *variabel* *warmup\_steps* bernilai 500 dengan *Regularisasi L2 weight\_decay* bernilai 0,01.

$$\text{lambda } x: 0 \text{ if } x \leq 2 \text{ else } 1 \text{ if } x == 3 \text{ else } 2 \quad (1)$$

Pengujian menggunakan data uji sebanyak 200. Metrik yang digunakan yaitu *precision*, *recall*, *F1-Score*, dan *accuracy*. *Precision* untuk mengukur seberapa akurat model dalam memprediksi kelas positif. *Recall* untuk mengukur seberapa baik model mendeteksi semua kelas positif. *F1-Score* (Persamaan 2) menggabungkan keseimbangan *precision* dan *recall*, sementara *accuracy* (Persamaan 3) mengukur proporsi prediksi benar terhadap total data sebagai indikator efektivitas global model.

$$F1Score = \frac{Precision \cdot Recall}{Precision + Recall} \quad (2)$$

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

**Keterangan:**

*True Positives (TP)* : Data yang diprediksi benar sebagai kelas tertentu.

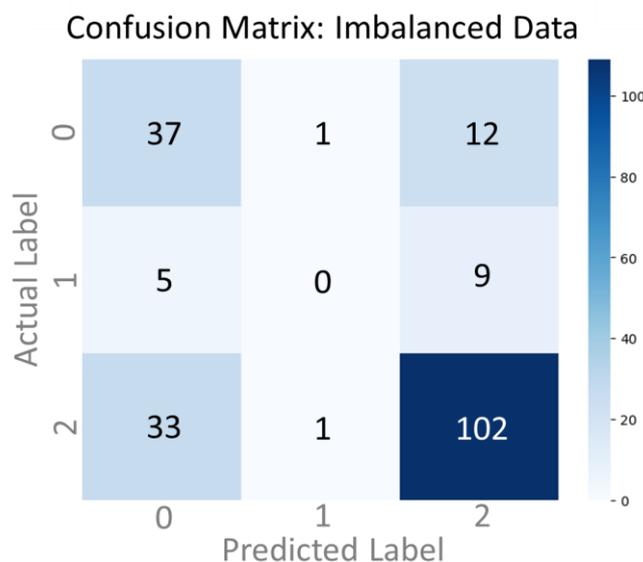
*True Negatives (TN)* : Data yang diprediksi benar bukan kelas tertentu.

*False Positives (FP)* : Data yang salah diprediksi sebagai kelas tertentu.

*False Negatives (FN)* : Data yang seharusnya kelas tertentu tetapi tidak terdeteksi.

**HASIL DAN PEMBAHASAN****Hasil**

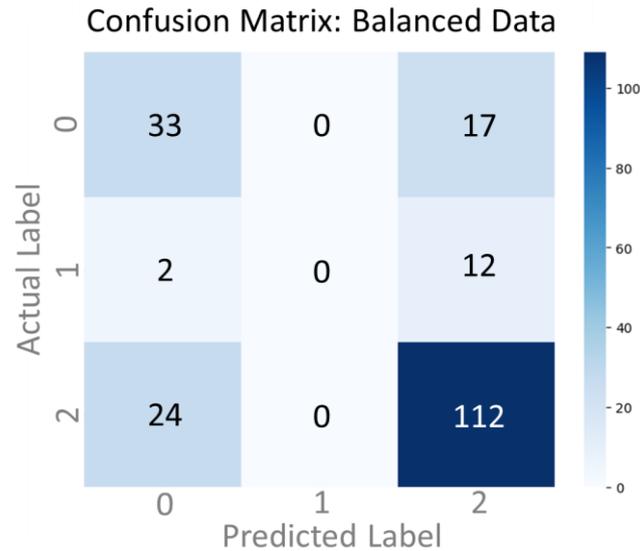
Evaluasi perbandingan antara data tidak seimbang dan data yang telah diolah dengan SMOTE untuk mencapai keseimbangan ditampilkan melalui *confusion matrix*. Gambar 2 menunjukkan kinerja model dalam memprediksi label dibandingkan dengan label sebenarnya. Baris pertama (kelas 0 atau negatif), model berhasil memprediksi 37 data dengan benar sebagai negatif (*true negative*). Namun, ada 1 data yang salah dikategorikan sebagai netral, dan 12 data salah diprediksi sebagai positif. Baris kedua (kelas 1 atau netral), performa model terlihat lemah. Tidak ada satu pun data yang benar-benar diprediksi sebagai netral (*true neutral* = 0). Sebaliknya, dari data yang seharusnya netral, 5 data salah dikelompokkan sebagai negatif, dan 9 data lainnya salah masuk ke kategori positif. Ini menunjukkan bahwa model kesulitan dalam mengidentifikasi ulasan-ulasan bersifat netral. Sementara itu, pada baris ketiga (kelas 2 atau positif), model bekerja cukup baik. Dari keseluruhan data positif, 102 data berhasil diprediksi dengan benar sebagai positif (*true positive*). Meski demikian, masih ada beberapa kesalahan, seperti 33 data positif yang salah dinyatakan sebagai negatif, serta 1 data lagi yang salah diklasifikasikan ke kategori netral.



**Gambar 2.** *Confusion matrix* pengujian pada data asli (tidak seimbang)

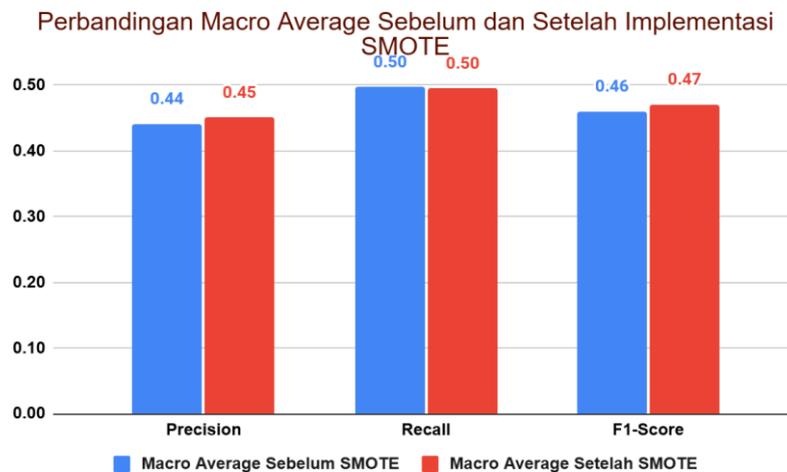
Hasil pada gambar 3 menunjukkan hasil pengujian setelah teknik SMOTE diterapkan dengan harapan dapat meningkatkan performa klasifikasi untuk Kelas 1. Pada kelas 0, model berhasil memprediksi 33 data dengan benar sebagai negatif (*true negatives*). Namun, ada 17 data yang salah dikategorikan sebagai positif, serta 2 data yang salah dikelompokkan sebagai netral. Pada kelas 1, model gagal sepenuhnya dalam memprediksi data netral karena tidak ada satu pun data yang benar-benar diklasifikasikan sebagai netral. Ada kesalahan klasifikasi dari netral ke negatif (2) dan positif (12). Kinerja pada kelas 2 menunjukkan model sangat baik dalam memprediksi data positif, dengan 112 data yang diprediksi dengan benar sebagai positif

(*true positives*). Namun, ada beberapa kesalahan klasifikasi ke kelas negatif (24), meskipun tidak ada kesalahan ke kelas netral.



**Gambar 3.** *Confusion matrix* pengujian pada data sintetis setelah smote (seimbang)

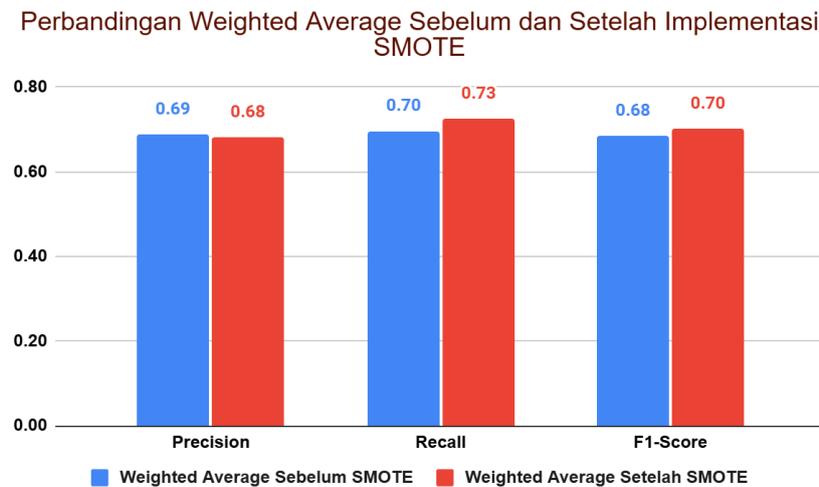
Hasil evaluasi menunjukkan bahwa performa untuk Kelas 1 tetap tidak membaik. Sebelum penerapan SMOTE, model tidak mampu mengenali satupun sampel dari Kelas 1, yang ditunjukkan oleh nilai TP sebesar 0. Setelah penerapan SMOTE, situasi ini masih belum berubah, karena model tetap gagal mengidentifikasi sampel dari Kelas 1. Dengan demikian, baik sebelum maupun sesudah penerapan SMOTE, performa model untuk Kelas 1 tidak mengalami perubahan signifikan.



**Gambar 4.** Perbandingan *macro average* sebelum dan setelah smote

Hasil metrik *macro average* pada gambar 4 menunjukkan kinerja model secara rata-rata untuk setiap kelas (tanpa mempertimbangkan proporsi kelas). Setelah SMOTE, peningkatan sedikit terjadi pada *precision* (0,45) dan *F1-Score* (0,47), tetapi *recall* tetap stagnan di 0,50. Meskipun SMOTE meningkatkan distribusi data dengan menciptakan sampel sintetis untuk kelas minoritas, dampaknya terhadap kinerja model (terutama pada Kelas 1) masih sangat terbatas. Selanjutnya pada hasil *weighted average* yang disajikan pada gambar 5 menunjukkan

bahwa model masih cenderung mengabaikan kelas minoritas, dan bias kelas tetap ada. Dengan metrik *weighted average*, hasil menunjukkan bahwa SMOTE memiliki dampak positif yang lebih terlihat pada kinerja agregat, terutama pada kelas mayoritas. Namun, karena *weighted average* lebih dipengaruhi oleh proporsi kelas, peningkatan ini tidak sepenuhnya mencerminkan perbaikan pada kelas minoritas. Model masih mengalami masalah dalam mendeteksi Kelas 1, yang tetap tidak terdeteksi sama sekali.



**Gambar 5.** Perbandingan *weighted average* sebelum dan setelah smote

### Pembahasan

Analisis kinerja IndoBERT dan data asli berdasarkan *macro average* menunjukkan beberapa hasil uji untuk dibahas. Skor *precision* (0,44) yang rendah menunjukkan bahwa model sering kali salah dalam memprediksi kelas minoritas sebagai positif palsu (*false positive*). Skor *recall* (0,50) sedikit lebih tinggi dari *precision* menunjukkan bahwa model berhasil mengidentifikasi sebagian besar sampel kelas minoritas, tetapi masih banyak yang terlewat (*false negative*). *F1-Score* (0,46) bernilai relatif rendah, yang menunjukkan bahwa model belum optimal dalam menangani kelas minoritas. Nilai-nilai ini relatif rendah, yang menunjukkan bahwa model kesulitan memprediksi kelas-kelas minoritas dengan baik. Hal ini umum terjadi pada set data yang tidak seimbang karena model cenderung bias ke kelas mayoritas.

Analisis kinerja IndoBERT dan data asli berdasarkan *weighted average* menunjukkan skor yang relatif lebih tinggi, yaitu skor *precision* (0,69), *recall* (0,70), dan *F1-Score* (0,68). Nilai-nilai ini jauh lebih tinggi dibandingkan *macro average*, yang menunjukkan bahwa model bekerja cukup baik pada kelas mayoritas (kelas dengan jumlah sampel lebih besar). Namun, performa pada kelas minoritas tetap buruk, seperti tercermin dari nilai *macro average* yang rendah. Model cenderung lebih fokus mempelajari pola dari kelas mayoritas karena representasi data mereka lebih dominan dalam *dataset*. Akibatnya, model kurang sensitif terhadap pola unik dari kelas minoritas, sehingga performa pada kelas tersebut menjadi lebih rendah.

SMOTE digunakan untuk meningkatkan jumlah ulasan pada kelas minoritas, sehingga distribusi jumlah ulasan di antara semua kelas menjadi lebih seimbang. Setelah penerapan SMOTE, setiap kelas memiliki 501 ulasan, menunjukkan bahwa teknik ini berhasil menambah data sintesis untuk kelas yang kurang terwakili dalam dataset asli. Analisis kinerja IndoBERT dan data sintesis berdasarkan *macro average* menunjukkan bahwa performa model secara keseluruhan masih cukup rendah, dengan nilai rata-rata sekitar 0,47. Hal ini mungkin

disebabkan oleh keterbatasan model dalam membedakan antara kelas minoritas dan mayoritas meskipun sudah diterapkan SMOTE.

Berdasarkan hasil pengujian, model memiliki performa yang lebih baik pada kelas mayoritas dibandingkan kelas minoritas. Hal ini tercermin dari perbedaan signifikan antara *macro average* (yang lebih rendah) dan *weighted average* (yang lebih tinggi). Sementara itu, *balanced accuracy* memberikan gambaran bahwa model cukup seimbang dalam memprediksi kedua kelas, tetapi masih ada ruang untuk perbaikan.

Pemanfaatan SMOTE pada untuk mengolah data numerik berhasil meningkatkan kinerja secara baik (Irfannandhy et al., 2024; Pardede & Pamungkas, 2024; Riandhoko et al., 2024). Penelitian-penelitian sebelumnya berfokus pada data numerik, di mana SMOTE bekerja dengan baik karena pola hubungan antarfitur lebih mudah dipetakan secara matematis. Data numerik biasanya memiliki struktur yang lebih homogen sehingga sintesis data minoritas cenderung lebih representatif. Namun, hasil pengujian pada penelitian ini (berbasis teks) menunjukkan kinerja yang kurang bagus karena memiliki kompleksitas lebih tinggi dibandingkan data numerik. Representasi data teks melibatkan dimensi yang jauh lebih besar (misalnya, vektor embedding dari kata-kata atau kalimat), dan pola bahasa sering kali tidak linier serta sulit direplikasi secara sintesis. Ini membuat SMOTE kurang efektif dalam menangani data teks, terutama jika data minoritas memiliki variasi linguistik yang kompleks (seperti sarkasme, dialek, atau frasa ambigu). Di sisi lain, kelas minoritas seringkali mengandung *noise* (e.g., *typo*, kalimat tidak baku) yang tidak tertangkap oleh SMOTE, sehingga model tetap gagal mempelajari pola yang relevan.

## SIMPULAN

Penelitian ini menemukan bahwa model *IndoBERT-base-uncased* mengalami kendala ketidakseimbangan kelas, dengan kecenderungan memprediksi kelas mayoritas (kelas 2) sebelum penerapan SMOTE. Meskipun SMOTE meningkatkan distribusi data dan metrik agregat seperti *precision*, *recall*, dan *F1-score*, peningkatan pada kelas minoritas (kelas 1) masih terbatas, menunjukkan persistensi bias kelas. Akurasi yang meningkat dari 69,5% menjadi 72,5% tidak sepenuhnya mencerminkan perbaikan kualitas model karena dominasi kelas mayoritas, sementara *F1-score* lebih representatif untuk evaluasi data tidak seimbang. Implikasi praktisnya, hasil ini dapat membantu pengembang aplikasi *game* mendeteksi masalah kritis dari ulasan bintang 1 untuk respons cepat dan strategi retensi pengguna. Untuk penelitian lanjutan, disarankan eksplorasi teknik *oversampling* adaptif (GAN), model *hybrid* (IndoBERT dengan attention/BiLSTM), atau kombinasi *undersampling* dan *ensemble learning*, serta integrasi fitur eksternal untuk memperkaya analisis sentimen.

## REFERENSI

- A'la, F. Y. (2022). Indonesian Sentiment Analysis towards MyPertamina Application Reviews by Utilizing Machine Learning Algorithms. *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, 5(1), 80–91. <https://doi.org/https://doi.org/10.20895/inista.v5i1.838>
- A'la, F. Y., Firdaus, N., Hartatik, & Safi'ie, M. A. (2023). SMOTE on Numeric Breast Cancer Dataset to Overcome Imbalance Class. *2023 6th International Conference of Computer and Informatics Engineering (IC2IE)*, 335–339. Indonesia: IEEE. <https://doi.org/10.1109/IC2IE60547.2023.10331221>
- Cahyono, B. E. H., & Sawitri, D. A. (2024). Pemerolehan Aspek Leksikon dan Struktur Kalimat Bahasa Indonesia: Sebuah Studi Kasus Anak Naya di Ponorogo. *Jurnal Onoma: Pendidikan, Bahasa, Dan Sastra*, 10(3), 3144–3162. <https://doi.org/10.30605/onoma.v10i3.4062>

- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Dharmendra, I. K., Putra, I. M. A. W., & Atmojo, Y. P. (2024). Evaluation of the Effectiveness of SMOTE and Random Under Sampling in Emotion Classification of Tweets. *Informatics for Educators and Professional: Journal of Informatics*, 9(2), 182-193. <https://doi.org/10.51211/itbi.v9i2.3183>
- Hidayat, W., Ardiansyah, M., & Setyanto, A. (2021). Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb. *Edumatic: Jurnal Pendidikan Informatika*, 5(1), 11–20. <https://doi.org/10.29408/edumatic.v5i1.3125>
- Imaduddin, H., A'la, F. Y., & Nugroho, Y. S. (2023). Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach. *International Journal of Advanced Computer Science and Applications*, 14(8)113-117. <https://doi.org/10.14569/IJACSA.2023.0140813>
- Imaduddin, H., Kusumaningtias, L. A., & A'la, F. Y. (2023). Application of LSTM and GloVe Word Embedding for Hate Speech Detection in Indonesian Twitter Data. *Ingénierie Des Systèmes d'Information*, 28(4), 1107–1112. <https://doi.org/10.18280/isi.280430>
- Irfannandhy, R., Handoko, L. B., & Ariyanto, N. (2024). Analisis Performa Model Random Forest dan CatBoost dengan Teknik SMOTE dalam Prediksi Risiko Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 714–723. <https://doi.org/10.29408/edumatic.v8i2.27990>
- Jimoh, I. A., Ismaila, I., & Olalere, M. (2019). Enhanced Decision Tree-J48 With SMOTE Machine Learning Algorithm for Effective Botnet Detection in Imbalance Dataset. *The 15th International Conference on Electronics, Computer and Computation (ICECCO)*, 1–8. Abuja, Nigeria: IEEE. <https://doi.org/10.1109/ICECCO48375.2019.9043233>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. *arXiv preprint arXiv:2011.00677*. <https://doi.org/10.18653/v1/2020.coling-main.66>
- Lv, D., Ma, Z., Yang, S., Li, X., Ma, Z., & Jiang, F. (2018). The Application of SMOTE Algorithm for Unbalanced Data. *Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality*, 10–13. Nagoya, Japan: ACM Digital Library. <https://doi.org/10.1145/3293663.3293686>
- Mohammed, A. J. (2020). Improving Classification Performance for a Novel Imbalanced Medical Dataset using SMOTE Method. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3), 3161–3172. <https://doi.org/10.30534/ijatcse/2020/104932020>
- Nuradita, R. D. (2022). Struktur Kalimat Dasar Bahasa Indonesia dalam Materi Aplikasi Duolingo (Kajian Sintaksis). *Wicara: Jurnal Sastra, Bahasa, Dan Budaya*, 1(2), 84–89.
- Pamungkas, A. A., Alam, C. N., Atmadja, A. R., & Juliansyah, R. (2024). Integrasi Kamus Multibahasa pada Feed Forward Neural Network dan IndoBERT dalam Pengembangan Chatbot Mobile. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 635–644. <https://doi.org/10.29408/edumatic.v8i2.27886>
- Pardede, J., & Pamungkas, D. P. (2024). The Impact of Balanced Data Techniques on Classification Model Performance. *Scientific Journal of Informatics*, 11(2), 401-412. <https://doi.org/10.15294/sji.v11i2.3649>
- Pradnyana, G. A., Anggraeni, W., Yuniarno, E. M., & Purnomo, M. H. (2023). Fine-Tuning IndoBERT Model for Big Five Personality Prediction from Indonesian Social Media. *International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 93–98. Bandung, Indonesia: IEEE. <https://doi.org/10.1109/ICoDSA58501.2023.10277572>

- Putra, A. B. Y. A., Sibaroni, Y., & Ihsan, A. F. (2023). Disinformation Detection on 2024 Indonesia Presidential Election using IndoBERT. *International Conference on Data Science and Its Applications (ICoDSA)*, 350–355. Bandung, Indonesia: IEEE. <https://doi.org/10.1109/ICoDSA58501.2023.10277572>
- Rahmawati, A., Alamsyah, A., & Romadhony, A. (2022). Hoax News Detection Analysis using IndoBERT Deep Learning Methodology. *2022 10th International Conference on Information and Communication Technology (ICoICT)*, 368–373. Bandung, Indonesia: IEEE. <https://doi.org/10.1109/ICoICT55009.2022.9914902>
- Riandhoko, A., Amalita, N., Vionanda, D., & Salma, A. (2024). Handling Unbalanced Data with SMOTE Algorithm for Unemployment Classification in Lima Puluh Kota Regency Using CART Method. *Indonesian Journal of Statistics and Its Applications*, 8(2), 166–177. <https://doi.org/10.29244/ijisa.v8i2p166-177>
- Setiawan, A., & Suryono, R. R. (2024). Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 183–192. <https://doi.org/10.29408/edumatic.v8i1.25667>
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., ... & Purwarianti, A. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. *arXiv preprint arXiv:2009.05387*. <https://doi.org/10.18653/v1/2020.aacl-main.85>
- Yulfa, R. I., Setiawan, B. H., Lourensius, G. G., & Purwandari, K. (2023). Enhancing Hate Speech Detection in Social Media Using IndoBERT Model: A Study of Sentiment Analysis during the 2024 Indonesia Presidential Election. *International Conference on Computer and Applications (ICCA)*, 1–6. Cairo, Egypt: IEEE. <https://doi.org/10.1109/ICCA59364.2023.10401700>