



PERBANDINGAN PERFORMA RANDOM FOREST DAN GRADIENT BOOSTING DALAM PREDIKSI PADA DATASET CUSTOMER SHOPPING TRENDS

Ferdiana Putri¹, Dede Brahma Arianto²

¹Manajemen, Fakultas Ekonomi, Universitas Muhammadiyah Surakarta

²Informatika, Fakultas Sains dan Teknik, Universitas Faletehan

¹b100210362@student.ums.ac.id, ²dede.brahma2@gmail.com

Abstract

This study compares the performance of two machine learning algorithms, Random Forest and Gradient Boosting, in predicting product categories using the Customer Shopping Trends dataset. The dataset exhibits an imbalanced class distribution, prompting the use of oversampling techniques to improve the representation of minority classes. The analysis process involves data exploration, preprocessing, and model evaluation based on metrics such as accuracy, precision, recall, and F1-score. The results indicate that Random Forest provides more consistent and superior performance compared to Gradient Boosting, particularly in handling minority classes. The Random Forest model achieved higher accuracy and more balanced evaluation metrics across all classes. This study offers insights into the effectiveness of ensemble algorithms in addressing data imbalance and their relevance for practical applications in industries such as e-commerce and customer data analysis.

Keywords: *Random Forest, Gradient Boosting, Data imbalance, Oversampling, Model evaluation, Customer data analysis, e-Commerce.*

Abstrak

Penelitian ini membandingkan performa dua algoritma pembelajaran mesin, yaitu *Random Forest* dan *Gradient Boosting*, dalam memprediksi kategori produk pada *dataset Customer Shopping Trends*. Dataset ini memiliki distribusi kelas yang tidak merata, sehingga teknik *oversampling* digunakan untuk meningkatkan representasi kelas minoritas. Proses analisis melibatkan eksplorasi data, pra-pemrosesan, dan evaluasi model berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa *Random Forest* memberikan performa yang lebih konsisten dan unggul dibandingkan *Gradient Boosting*, terutama dalam menangani kelas minoritas. Model *Random Forest* berhasil mencapai akurasi yang lebih tinggi dan nilai metrik evaluasi yang lebih seimbang pada seluruh kelas. Penelitian ini memberikan wawasan tentang efektivitas algoritma *ensemble* dalam menghadapi ketidakseimbangan data, serta relevansinya untuk aplikasi praktis di industri, seperti *e-commerce* dan analisis data pelanggan.

Kata Kunci: *Random Forest, Gradient Boosting, Ketidakseimbangan data, Oversampling, Evaluasi model, Analisis data pelanggan, e-Commerce*

Article History

Received: December 2024

Reviewed: December 2024

Published: December 2024

Plagiarism Checker No 234

Prefix DOI :

10.8734/Kohesi.v1i2.365

Copyright : Author

Publish by : Kohesi



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



PENDAHULUAN

Dalam era digital yang serba cepat, data menjadi salah satu aset terpenting dalam pengambilan keputusan. Data yang akurat dan relevan dapat membantu bisnis memahami pola perilaku pelanggan, memprediksi kebutuhan pasar, serta merancang strategi pemasaran yang lebih efektif. Pembelajaran mesin telah menjadi alat yang sangat berharga untuk menganalisis dan memanfaatkan data. Pada pembelajaran mesin, algoritma mampu dimanfaatkan dalam mengidentifikasi pola dalam data dan membuat prediksi berdasarkan pola yang ditunjukkan (Oktavianus et al., 2023). Algoritma pembelajaran mesin memungkinkan komputer mengenali pola dari data besar dan kompleks serta membuat prediksi dengan tingkat akurasi yang tinggi. Namun, efektivitas algoritma sangat bergantung pada kualitas dan distribusi data yang digunakan.

Ketidakseimbangan data atau *imbalanced dataset* adalah salah satu tantangan utama yang sering dihadapi. Ketidakseimbangan terjadi ketika jumlah sampel dalam satu atau lebih kelas jauh lebih sedikit dibandingkan dengan kelas lainnya. Sebagai contoh, dalam *dataset "Customer Shopping Trends,"* kategori "*Clothing*" memiliki jumlah data yang jauh lebih banyak dibandingkan kategori "*Accessories*." Ketidakseimbangan seperti ini dapat menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga menghasilkan prediksi yang kurang akurat untuk kelas minoritas. Situasi ini menjadi lebih kritis dalam aplikasi praktis, di mana kelas minoritas sering kali memiliki kepentingan bisnis yang tinggi, seperti mendeteksi pelanggan yang membeli produk premium atau memahami kebutuhan segmen pasar yang kecil.

Dua algoritma pembelajaran mesin yang sering digunakan untuk menangani masalah klasifikasi adalah *Random Forest* dan *Gradient Boosting*. *Random Forest* adalah algoritma berbasis *ensemble* yang menggabungkan hasil dari beberapa pohon keputusan untuk meningkatkan akurasi dan stabilitas prediksi. Algoritma ini dikenal karena kemampuannya dalam menangani *dataset* dengan banyak fitur serta toleransi terhadap data yang berisik atau mengandung *outlier*. Di sisi lain, *Gradient Boosting* bekerja dengan cara membangun model secara iteratif, di mana setiap iterasi berfokus pada memperbaiki kesalahan prediksi dari model sebelumnya. Algoritma ini memiliki kemampuan untuk memberikan performa tinggi, terutama ketika diterapkan pada data yang bersih dan diproses dengan baik. Namun, *Gradient Boosting* cenderung lebih sensitif terhadap ketidakseimbangan data dibandingkan *Random Forest*.

Analisis dilakukan untuk membandingkan performa kedua algoritma tersebut pada *dataset "Customer Shopping Trends."* Proses analisis mencakup beberapa tahapan penting, mulai dari eksplorasi data, deteksi dan penanganan *outlier*, penanganan ketidakseimbangan data dengan teknik *oversampling*, hingga evaluasi model menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-score*. Dengan memanfaatkan *dataset* ini, tujuan utama dari analisis adalah untuk memahami algoritma mana yang lebih unggul dalam menangani ketidakseimbangan data serta bagaimana teknik pra-pemrosesan data dapat memengaruhi hasil prediksi.

Analisis *dataset "Customer Shopping Trends"* relevan karena menggambarkan situasi nyata dalam dunia ritel, di mana transaksi pelanggan sering kali memiliki distribusi data yang tidak merata. Misalnya, kategori produk yang lebih populer seperti "*Clothing*" sering kali mendominasi, sedangkan kategori yang lebih *niche* seperti "*Accessories*" hanya memiliki sebagian kecil data. Dengan latar belakang ini, penting untuk memahami algoritma mana yang lebih efektif dalam



menangani *dataset* tidak seimbang, terutama dalam konteks penerapan praktis seperti rekomendasi produk, analisis pasar, dan optimisasi strategi penjualan.

TINJAUAN LITERATUR

Ketidakseimbangan Data

Ketidakseimbangan data adalah salah satu tantangan utama dalam pembelajaran mesin yang memengaruhi akurasi dan efektivitas model prediksi. Kondisi ini terjadi ketika distribusi kelas dalam *dataset* tidak merata, di mana satu kelas memiliki jumlah sampel yang jauh lebih banyak dibandingkan kelas lainnya. Ketidakseimbangan ini sering kali menyebabkan model menjadi bias terhadap kelas mayoritas, sehingga performanya untuk kelas minoritas sangat rendah (Meidianingsih & Agustine, 2021). Dalam domain seperti analisis data pelanggan, deteksi penipuan, atau diagnosis medis, keberhasilan model dalam mengenali kelas minoritas sangatlah penting karena kelas ini sering kali memiliki nilai bisnis atau implikasi klinis yang signifikan. Misalnya, dalam prediksi kategori pelanggan, kelas minoritas mungkin mencakup pelanggan dengan nilai *lifetime* yang tinggi yang harus ditargetkan oleh strategi pemasaran perusahaan.

Salah satu dampak utama dari ketidakseimbangan data adalah rendahnya performa model dalam metrik seperti *precision*, *recall*, dan *F1-score* untuk kelas minoritas (Kar et al., 2022). Akurasi keseluruhan sering kali memberikan gambaran yang menyesatkan karena model dapat mencapai skor akurasi tinggi dengan hanya memprediksi kelas mayoritas. Oleh karena itu, diperlukan pendekatan khusus untuk mengatasi masalah ini. Teknik seperti *oversampling* telah banyak digunakan untuk menyeimbangkan distribusi data, dengan fokus pada peningkatan representasi kelas minoritas tanpa mengubah distribusi fitur lainnya (Hameed & Bawany, 2022). Selain itu, modifikasi algoritma dan penggunaan metrik evaluasi yang lebih spesifik juga menjadi bagian penting dari strategi mitigasi (Johnson & Khoshgoftaar, 2019).

Teknik Penanganan Ketidakseimbangan Data

Teknik penanganan ketidakseimbangan data dapat dibagi menjadi tiga kategori utama: metode berbasis data, metode berbasis algoritma, dan metode berbasis *hybrid*. Metode berbasis data melibatkan manipulasi *dataset* sebelum pelatihan model, termasuk *oversampling*, *undersampling*, dan SMOTE (*Synthetic Minority Over-sampling Technique*). SMOTE adalah teknik yang menggunakan interpolasi untuk menghasilkan sampel sintesis dari data minoritas yang ada, sehingga meningkatkan jumlah data kelas minoritas tanpa mengurangi keragaman data (Development et al., 2019).

Metode berbasis algoritma mencakup modifikasi algoritma pembelajaran untuk menangani ketidakseimbangan data secara langsung. Contohnya adalah penyesuaian bobot kelas dalam algoritma *Gradient Boosting* dan *Random Forest* untuk memberi penalti lebih tinggi pada kesalahan prediksi kelas minoritas. Metode *hybrid* menggabungkan manipulasi data dengan algoritma yang disesuaikan untuk memaksimalkan performa model. Pendekatan ini sering kali digunakan dalam aplikasi dunia nyata, di mana kompleksitas *dataset* membutuhkan solusi yang lebih fleksibel (Development et al., 2019).



Random Forest:

Random Forest adalah algoritma *ensemble* yang menggabungkan prediksi dari beberapa pohon keputusan independen. Pendekatan ini membuat *Random Forest* tahan terhadap *overfitting* dan mampu menangani *dataset* dengan banyak fitur. Algoritma ini sering kali digunakan untuk menangani ketidakseimbangan data dengan fitur *bagging* yang membantu meminimalkan bias terhadap kelas mayoritas. *Random Forest* juga dikenal karena efisiensinya dalam menangani data berukuran besar dan kompleks (Development et al., 2019).

Gradient Boosting:

Gradient Boosting adalah algoritma *ensemble* berbasis boosting yang bekerja secara iteratif untuk memperbaiki kesalahan prediksi dari model sebelumnya. Algoritma ini menghasilkan model dengan akurasi tinggi tetapi cenderung lebih sensitif terhadap ketidakseimbangan data (Dorogush et al., n.d.).

Metrik Evaluasi

Dalam kasus ketidakseimbangan data, metrik seperti *accuracy* kurang representatif karena tidak memberikan gambaran yang jelas tentang performa model untuk kelas minoritas. *Precision* mengukur akurasi prediksi pada kelas positif, *recall* mengukur kemampuan model untuk mendeteksi kelas positif secara keseluruhan, dan *F1-score* memberikan keseimbangan antara *precision* dan *recall*. Metrik-metrik ini sangat penting dalam mengevaluasi performa model dalam konteks ketidakseimbangan data karena memberikan fokus pada kemampuan model untuk mengenali kelas minoritas (Johnson & Khoshgoftaar, 2019).

Relevansi dalam Domain E-commerce

Semakin meningkatnya *e-commerce* saat ini telah memungkinkan penjual untuk terhubung dengan pelanggan potensial dengan cara yang baru dan memungkinkan memudahkan dalam melakukan analisis (Liu et al., 2022). Dalam domain *e-commerce*, analisis data pelanggan sering kali dihadapkan pada ketidakseimbangan data, misalnya dalam memprediksi preferensi pelanggan untuk kategori produk tertentu. Penelitian oleh (Med, 2020) menunjukkan bahwa kombinasi algoritma *ensemble* seperti *Random Forest* dengan teknik *oversampling* secara signifikan meningkatkan performa prediksi. Selain itu, pentingnya memilih metrik evaluasi yang sesuai untuk memastikan keberhasilan implementasi model dalam skenario dunia nyata.

METODOLOGI

Eksplorasi Data

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, dengan judul "*Customer Shopping Latest Trends Dataset*" yang disediakan oleh Bhadra Mohit (<https://www.kaggle.com/datasets/bhadramohit/customer-shopping-latest-trends-dataset>).

Dataset ini berisi informasi tentang tren belanja pelanggan, mencakup atribut seperti usia pelanggan, jumlah pembelian, dan kategori produk. Analisis dimulai dengan pemeriksaan struktur *dataset* menggunakan *library* *pandas*, termasuk jumlah kolom, tipe data, serta nilai yang hilang.

Deteksi dan Penanganan outlier:

Outlier pada kolom numerik dideteksi menggunakan metode *Z-score* dan *IQR (Interquartile Range)*. *Z-score* menghitung deviasi nilai terhadap rata-rata dalam satuan standar deviasi, sementara *IQR* menggunakan kuartil untuk mengidentifikasi nilai yang ekstrem. Nilai-nilai *outlier* yang terdeteksi diganti dengan rata-rata kolom untuk menjaga stabilitas data. Hal ini



penting karena *outlier* dapat menyebabkan bias pada algoritma pembelajaran mesin, terutama yang sensitif terhadap skala data. Penanganan *outlier* terbukti meningkatkan stabilitas model (Gans, 2019).

Penanganan Ketidakseimbangan Data:

Ketidakseimbangan data pada kategori target ditangani menggunakan *Random Oversampler* dari *library imbalanced-learn*. Teknik ini meningkatkan jumlah data pada kelas minoritas hingga setara dengan kelas mayoritas. *Oversampling* dilakukan setelah *dataset* dibagi menjadi fitur (X) dan target (y). Metode ini dipilih karena sederhana namun efektif dalam meningkatkan representasi kelas minoritas tanpa mengubah distribusi fitur lain. Hasil distribusi kategori setelah *oversampling* divisualisasikan menggunakan grafik batang.

Pemodelan:

1. *Gradient Boosting*:

Algoritma ini digunakan dengan parameter *default* untuk mengevaluasi performa awal. *Gradient Boosting* membangun model secara interaktif dengan menambahkan model baru yang fokus pada kesalahan prediksi sebelumnya. Dalam implementasi, digunakan *GradientBoostingClassifier* dari *library scikit-learn*.

2. *Random Forest*:

Sama seperti *Gradient Boosting*, algoritma ini digunakan dengan parameter *default*. *Random Forest* menghasilkan beberapa pohon keputusan independen dan menggabungkan hasil prediksi mereka untuk menghasilkan prediksi akhir yang lebih stabil.

Evaluasi

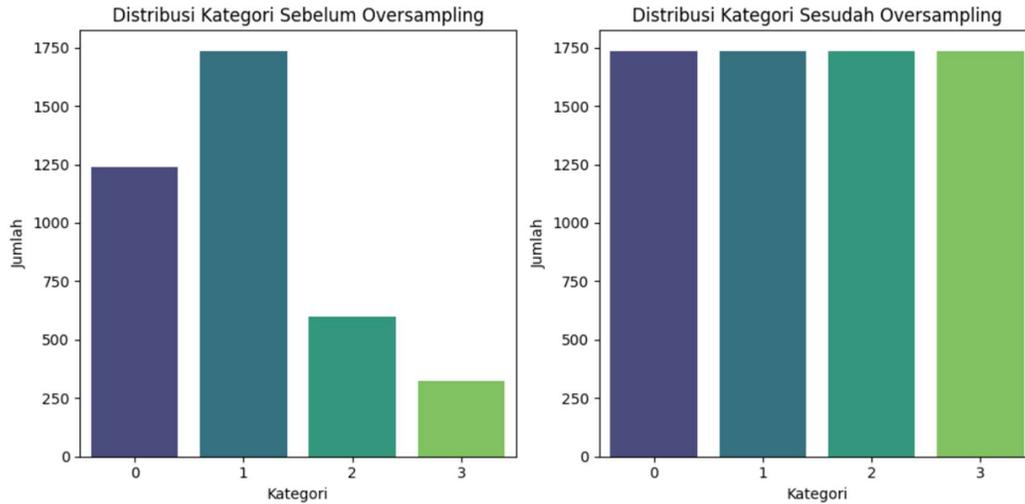
Evaluasi dilakukan menggunakan *Classification Report*. *Classification Report* merangkum metrik seperti *precision*, *recall*, dan *F1-score* (Marpaung et al., 2024). Hasil dari evaluasi dapat digunakan untuk melihat akurasi dari algoritma yang digunakan.

HASIL DAN PEMBAHASAN

Analisis awal menunjukkan bahwa distribusi kategori pada *dataset "Customer Shopping Trends"* sangat tidak seimbang. Kategori "*Clothing*" mendominasi dengan jumlah data yang jauh lebih besar dibandingkan kategori lainnya seperti "*Accessories*." Ketidakseimbangan ini dapat menyebabkan bias model terhadap kelas mayoritas, sehingga diperlukan teknik penyeimbangan data. *Oversampling* menggunakan metode *Random Oversampler* diterapkan untuk meningkatkan jumlah data pada kelas minoritas tanpa mengubah distribusi fitur lainnya. Setelah *oversampling*, distribusi kategori menjadi seimbang, yang diharapkan dapat meningkatkan kemampuan model dalam mengenali pola pada kelas minoritas. Visualisasi distribusi kategori sebelum dan sesudah *oversampling* menunjukkan bahwa metode ini berhasil meningkatkan representasi kelas minoritas secara signifikan.



Gambar 1. Distribusi Kategori Sebelum dan Sesudah *Oversampling*



Hasil penelitian menunjukkan perbedaan yang signifikan dalam performa antara *Gradient Boosting* dan *Random Forest* pada *dataset Customer Shopping Trends*. Kedua model dievaluasi menggunakan konfigurasi *default* untuk menilai efektivitas algoritma secara langsung. Meskipun kedua algoritma mampu menangani data dengan baik, performa *Gradient Boosting* terlihat kurang optimal dibandingkan *Random Forest*, terutama dalam menghadapi kelas minoritas yang jumlahnya jauh lebih sedikit dibandingkan kelas mayoritas. Hal ini mengindikasikan pentingnya memahami karakteristik algoritma serta dampaknya pada distribusi data yang tidak seimbang.

Pada *Gradient Boosting Classifier*, akurasi model mencapai 42%, yang secara umum menunjukkan hasil kurang memuaskan. *Precision* rata-rata (*macro average*) untuk model ini hanya mencapai 0.26, yang berarti sebagian besar prediksi untuk kelas minoritas tidak akurat. Sebagai contoh, *precision* untuk kelas *Accessories* (kelas 3) adalah 0.00, menunjukkan bahwa model ini tidak mampu memprediksi kategori tersebut dengan benar. *Recall* untuk kelas minoritas juga sangat rendah, dengan nilai 0.02 untuk kelas *Outerwear* (kelas 2) dan 0.00 untuk kelas *Accessories*. Hal ini menunjukkan bahwa *Gradient Boosting* cenderung mengabaikan kelas dengan jumlah data yang lebih sedikit, yang konsisten dengan sensitivitasnya terhadap distribusi data yang tidak merata. *F1-score* pada *Gradient Boosting* mempertegas ketidakmampuan model dalam menangkap pola dari kelas minoritas. Nilai *F1-score* untuk kelas *Accessories* adalah 0.00, sementara untuk kelas mayoritas seperti *Clothing* (kelas 0) mencapai 0.20. Nilai ini menunjukkan bahwa model lebih efektif dalam memprediksi kelas mayoritas dibandingkan kelas minoritas.

Gambar 2. *Classification Report Gradient Boosting*

Classification Report:				
	precision	recall	f1-score	support
0	0.32	0.15	0.20	249
1	0.44	0.84	0.58	346
2	0.29	0.02	0.03	122
3	0.00	0.00	0.00	63
accuracy			0.42	780
macro avg	0.26	0.25	0.20	780
weighted avg	0.34	0.42	0.33	780



Sebaliknya, *Random Forest Classifier* menunjukkan performa yang jauh lebih baik dengan akurasi mencapai 85%. Model ini unggul dalam menangkap pola dari semua kelas, termasuk kelas minoritas. *Precision* rata-rata (*macro average*) mencapai 0.82, dengan *precision* untuk kelas *Accessories* mencapai 0.71. Hal ini menunjukkan bahwa *Random Forest* mampu memberikan prediksi yang lebih akurat untuk kelas minoritas. *Recall* untuk kelas *Accessories* mencapai 1.00, menunjukkan bahwa semua data dalam kelas ini berhasil diprediksi dengan benar oleh model. Keunggulan *Random Forest* terletak pada mekanisme *ensemble* berbasis *bagging* yang memanfaatkan banyak pohon keputusan untuk memberikan prediksi yang lebih stabil dan andal.

F1-score untuk *Random Forest* juga konsisten lebih tinggi dibandingkan *Gradient Boosting*. Rata-rata F1-score mencapai 0.85, yang mencerminkan keseimbangan yang baik antara *precision* dan *recall*. Model ini menunjukkan nilai F1-score yang tinggi untuk semua kelas, termasuk kelas mayoritas dan minoritas, yang menegaskan stabilitasnya dalam menangani *dataset* dengan distribusi data yang tidak seimbang. Dengan menggunakan teknik *oversampling* untuk meningkatkan representasi kelas minoritas, *Random Forest* dapat memaksimalkan potensi algoritma *ensemble* untuk meningkatkan performa prediksi secara keseluruhan.

Gambar 3. *Classification Report Random Forest*

Classification Report:				
	precision	recall	f1-score	support
0	0.84	0.82	0.83	249
1	0.91	0.81	0.86	346
2	0.81	0.93	0.87	122
3	0.71	1.00	0.83	63
accuracy			0.85	780
macro avg	0.82	0.89	0.85	780
weighted avg	0.86	0.85	0.85	780

Perbandingan antara kedua algoritma ini juga mencerminkan sifat dasar masing-masing model. *Gradient Boosting*, meskipun efektif dalam menangani kesalahan iteratif, lebih sensitif terhadap distribusi data yang tidak merata. Di sisi lain, *Random Forest* lebih *robust* terhadap ketidakseimbangan data karena setiap pohon keputusan dalam *ensemble*-nya berkontribusi secara independen terhadap prediksi akhir. Hal ini memungkinkan model untuk memberikan hasil yang lebih stabil meskipun kelas minoritas memiliki jumlah sampel yang kecil.

Hasil ini juga sesuai dengan temuan dari penelitian sebelumnya yang menunjukkan bahwa *Random Forest* sering kali lebih efektif dibandingkan algoritma *ensemble* lainnya dalam menangani *dataset* yang tidak seimbang (Development et al., 2019). Keunggulan ini terutama terlihat dalam aplikasi yang melibatkan data kategori dengan distribusi yang tidak merata, seperti analisis pelanggan dalam *e-commerce*. Penelitian ini juga menegaskan pentingnya teknik pra-pemrosesan data seperti *oversampling*, yang terbukti meningkatkan performa model dalam memprediksi kelas minoritas.

Secara keseluruhan, penelitian ini memberikan wawasan penting tentang bagaimana algoritma pembelajaran mesin bereaksi terhadap *dataset* dengan ketidakseimbangan data. *Random Forest* terbukti menjadi pilihan yang lebih andal dan praktis dibandingkan *Gradient Boosting* dalam skenario ini. Implikasi dari penelitian ini sangat relevan untuk aplikasi praktis, seperti sistem rekomendasi produk atau segmentasi pelanggan, di mana prediksi yang akurat



untuk kelas minoritas sering kali memiliki dampak yang signifikan terhadap pengambilan keputusan bisnis.

KESIMPULAN

Hasil penelitian ini menunjukkan bahwa *Random Forest* memiliki keunggulan yang signifikan dibandingkan *Gradient Boosting* dalam memprediksi kategori produk pada *dataset Customer Shopping Trends*. Dengan pendekatan *ensemble* berbasis *bagging*, *Random Forest* berhasil memberikan hasil yang konsisten dan akurat pada semua kelas, termasuk kelas minoritas yang biasanya sulit diprediksi. Evaluasi metrik menunjukkan bahwa *Random Forest* unggul dalam *precision*, *recall*, dan *F1-score*, menjadikannya algoritma yang lebih andal untuk menangani dataset dengan ketidakseimbangan data.

Sebaliknya, *Gradient Boosting*, meskipun efektif dalam mendeteksi kelas mayoritas, menunjukkan kesulitan dalam memprediksi kelas minoritas, yang terlihat dari rendahnya nilai *recall* dan *F1-score* pada kelas tersebut. Hal ini mengindikasikan bahwa algoritma *Gradient Boosting* memerlukan perhatian lebih dalam penyesuaian atau penerapan teknik tambahan untuk meningkatkan performanya dalam menghadapi ketidakseimbangan data.

Secara keseluruhan, penelitian ini menegaskan pentingnya memilih algoritma yang sesuai dan menggunakan teknik pra-pemrosesan data seperti *oversampling* untuk memaksimalkan performa prediksi, khususnya pada data yang tidak seimbang. Hasil ini relevan untuk aplikasi dunia nyata, terutama dalam industri *e-commerce*, di mana segmentasi pelanggan dan analisis kategori produk sangat bergantung pada akurasi prediksi model.

Saran

1. Eksplorasi Algoritma Lain

Penelitian selanjutnya dapat mengeksplorasi algoritma lain, seperti *Bagging* atau *LightGBM*, yang dikenal memiliki performa tinggi pada data tidak seimbang. Kombinasi algoritma ini dengan teknik pra-pemrosesan yang lebih kompleks juga patut dipertimbangkan.

2. Validasi yang Lebih Luas

Disarankan untuk menggunakan validasi silang (*cross-validation*) dengan skema yang lebih kompleks guna memastikan generalisasi model ke *dataset* yang lebih besar atau dengan karakteristik berbeda.



DAFTAR PUSTAKA

- Development, S., Kauffman, I., Khai, L., Lien, T., & Jr, P. R. F. (2019). *Fernandez, P. R. 2018. Literature Review on Sustainability Science for HIGHLIGHTS OF A LITERATURE REVIEW ON SUSTAINABILITY SCIENCE FOR SUSTAINABLE DEVELOPMENT AND ITS IMPLICATIONS TO THE ASIA- PACIFIC REGION. September.*
- Dorogush, A. V., Ershov, V., & Gulin, A. (n.d.). *CatBoost: Gradient Boosting with categorical features support.* 1–7.
- Gans, J. S. (2019). *Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction.*
- Hameed, A., & Bawany, N. Z. (2022). *Network intrusion detection using oversampling technique and machine learning algorithms.* <https://doi.org/10.7717/peerj-cs.820>
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data.* <https://doi.org/10.1186/s40537-019-0192-5>
- Kar, A. K., Choudhary, S. K., & Singh, V. K. (2022). How can artificial intelligence impact sustainability: A systematic literature review. *Journal of Cleaner Production, 376*(May), 134120. <https://doi.org/10.1016/j.jclepro.2022.134120>
- Liu, X., Tong, D., Huang, J., Zheng, W., Kong, M., & Zhou, G. (2022). Land Use Policy What matters in the *e-commerce* era? Modelling and mapping shop rents in. *Land Use Policy, 123*(February), 106430. <https://doi.org/10.1016/j.landusepol.2022.106430>
- Marpaung, F., Khairina, N., & Muliono, R. (2024). *KLASIFIKASI DAUN TEH SIAP PANEN MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK ARSITEKTUR MOBILENETV2.* 18, 215–225.
- Med, J. T. (2020). Comments on: Huang et al. (2019) Emerging trends and research foci in gastrointestinal. *Journal of Translational Medicine, 1–2.* <https://doi.org/10.1186/s12967-020-02379-9>
- Meidianingsih, Q., & Agustine, D. (2021). *Study of Bagging Application in the Safe-Level Smote Method in Kajian Penerapan Bagging pada Metode Safe-Level Smote dalam.* 5(1), 105–116.
- Oktavianus, A. J. E., Naibaho, L., & Rantung, D. A. (2023). *Pemanfaatan Artificial Intelligence pada Pembelajaran dan Asesmen di Era Digitalisasi.* 05(2), 473–486.