RABIT : Jurnal Teknologi dan Sistem Informasi Univrab ISSN CETAK : 2477-2062
Volume 10 No. 1 | Januari 2025 : Hal : 51-59 ISSN ONLINE : 2502-891X

IMPLEMENTASI SMOTE-ENN DAN BORDERLINE SMOTE TERHADAP PERFORMA *LIGHTGBM* PADA IMBALANCED CLASS

[1)] Yabes Dwi Nugroho H, [2)] Furqan Zakiyabarsi, [3)] Andi Jamiati Paramita
[1,2,3)]Program Studi Sistem Informasi dan Teknologi, Institut Teknologi dan Bisnis Kalla
[1,2,3)]Makassar, Indonesia
E-mail : *dwi@kallainstitute.ac.id*, *furqan@kallainstitute.ac.id*, *andijeparamita@gmail.com*

ABSTRAK
Ketidakseimbangan kelas (*imbalanced class*) merupakan tantangan yang sering dihadapi dalam pengembangan model pembelajaran mesin, di mana distribusi data yang tidak merata antara kelas mayoritas dan kelas minoritas dapat menyebabkan bias prediksi terhadap kelas mayoritas. Penelitian ini mengkaji implementasi dua metode penyeimbangan data, yaitu SMOTE-ENN (*Synthetic Minority Over-sampling Technique and Edited Nearest Neighbor*) dan Borderline-SMOTE, untuk meningkatkan performa model LightGBM pada dataset *Online Shopperâ€™s Purchase Intention*. Dataset ini memiliki ketidakseimbangan distribusi antara kelas pembelian (*True*) dan non-pembelian (*False*), yang menghambat kemampuan model dalam mendeteksi kelas minoritas. Metode SMOTE-ENN mengombinasikan teknik *oversampling* untuk menciptakan sampel sintetik pada kelas minoritas dengan penghapusan sampel *noise* atau salah klasifikasi dari kelas mayoritas, sedangkan Borderline-SMOTE menghasilkan sampel sintetik di sekitar titik-titik kelas minoritas yang berada di dekat batas keputusan (*decision boundary*). Penelitian ini mengevaluasi performa model LightGBM sebelum dan sesudah penerapan kedua metode dengan menggunakan metrik evaluasi, seperti akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa kedua metode berhasil meningkatkan kemampuan model dalam mendeteksi kelas minoritas secara signifikan, dengan SMOTE-ENN memberikan keunggulan dalam menghasilkan distribusi data yang lebih representative dengan akurasi 93% dan Borderline-SMOTE 92%. Penelitian ini membuktikan efektivitas SMOTE-ENN dan Borderline-SMOTE dalam mengatasi ketidakseimbangan kelas

Kata Kunci: *Borderline SMOTE, Ketidakseimbangan Kelas, LightGBM., Purchase Intention, SMOTE-ENN*

ABSTRACT
Class imbalance is a significant challenge in machine learning, where unequal distribution between majority and minority classes often biases model predictions toward the majority class. This study investigates the implementation of two data balancing techniques, SMOTE-ENN (Synthetic Minority Over-sampling Technique and Edited Nearest Neighbor) and Borderline-SMOTE, to enhance the performance of the LightGBM model on the Online Shopperâ€™s Purchase Intention dataset. The dataset exhibits an imbalanced distribution between the purchase (True) and non-purchase (False) classes, hindering the modelâ€™s ability to detect minority classes accurately. The SMOTE-ENN method combines oversampling, which creates synthetic samples for the minority class, with noise removal by eliminating misclassified samples from the majority class. On the other hand, Borderline-SMOTE generates synthetic samples near the decision boundary of the minority class, focusing on critical regions prone to misclassification. The study evaluates the LightGBM modelâ€™s performance before and after applying these techniques using evaluation metrics such as accuracy, precision, recall, and F1-score. Results demonstrate that both methods significantly improve the modelâ€™s ability to detect the minority class, with Borderline-SMOTE showing a slight advantage by generating a more representative data distribution around the decision boundary. The results indicate that both methods significantly improve the modelâ€™s ability to detect the minority class, with SMOTE-ENN achieving an accuracy of 93% and demonstrating superiority in producing a more representative data distribution compared to Borderline-SMOTE, which achieved 92% accuracy. This study confirms the effectiveness of SMOTE-ENN and Borderline-SMOTE in addressing class imbalance in machine learning applications

Keyword: *Borderline SMOTE, Class Imbalance, LightGBM, Purchase Intention, SMOTE-ENN.*

DOI : https://doi.org/10.36341/rabit.v10i1.5436

PENDAHULUAN

Perkembangan teknologi atau transformasi digital disektor perdagangan seperti *e-commerce* telah meningkatkan popularitas belanja daring secara global dan mempengaruhi perilaku konsumen[1]. Salah satu alasan peningkatan dan perubuhan perilaku konsumen dipengaruhi oleh kemudahan dan variasi pilihan kepada konsumen[2]. Oleh karena itu, fenomena ini memotivasi penelitian tentang perilaku yang berfokus pada niat beli konsumen online yang menjadi indikator penting untuk memahami perilaku belanja secara digital. Untuk memahami perilaku diperlukan beberapa data terkait belanja konsumen secara digital misalnya analisis pola kunjungan dan sebagainya[3]. Dalam proses analisis sering terjadi permasalahan seperti dataset yang digunakan sering kali tidak seimbang. Apabila terdapat kelas yang tidak seimbang yaitu terdapat kelas mayoritas yang dianggap menarik oleh algoritma daripada kelas minoritas yang diabaikan dan dianggap sebagai *noise* karena tidak cukup terwakili. Ketidakseimbangan ini menjadi tantangan dalam penerapan algoritma karena menyebabkan algoritma cenderung bias terhadap kelas mayoritas sehingga sulit memberikan prediksi yang akurat.

Penelitian sebelumnya menunjukkan bahwa *Random Forest* memberikan akurasi yang tinggi dalam memprediksi niat pembelian konsumen, dengan hasil yang lebih baik dibandingkan algoritma lainnya seperti *Naive Bayes*[4]. Ahsain S memprediksi kemungkinan niat pembelian konsumen di platform *e-commerce* dengan menggunakan model pembelajaran mesin seperti *Decision Tree*, *Random Forest*, dan *Gradient Boosting*[5].

Tabel 1. Dataset Online Shopper Purchase Intention

No	Nama Fitur	Tipe Data	Deskripsi
1.	*Administrative*	Numerik	Jumlah halaman administrative yang dikunjungi selama sesi
2.	*Administrative_Duration*	Numerik	Total waktu (detik) yang dihabiskan pada halaman administratif.
3.	*Informational*	Numerik	Jumlah halaman informasi yang dikunjungi selama sesi.
4.	*Informational_Duration*	Numerik	Total waktu (detik) pada halaman informasi
5.	*ProductRelated*	Numerik	Jumlah halaman terkait produk yang dikunjungi selama sesi.
6.	*ProductRelated_Duration*	Numerik	Total waktu (dalam detik) yang dihabiskan pada halaman terkait produk
7.	*BounceRates*	Numerik	Persentase sesi yang hanya mengunjungi satu halaman sebelum meninggalkan situs
8.	*ExitRates*	Numerik	Persentase halaman yang menjadi halaman terakhir yang dilihat selama sesi
9.	*PageValues*	Numerik	Nilai rata-rata yang diberikan pada halaman berdasarkan konversi
10.	*SpecialDay*	Numerik	Indikator kedekatan dengan hari spesial (seperti Black Friday)
11.	*Month*	Kategorical	Bulan kunjungan (Jan, Feb, Mar, dll.)
12.	*OperatingSystems*	Kategorical	Sistem operasi pengguna (1, 2, 3, dll.)
13.	*Browser*	Kategorical	Jenis browser yang digunakan (1, 2, 3, dll.)

DOI : https://doi.org/10.36341/rabit.v10i1.5436 *Corresponding Author* : Yabes Dwi Nugroho H

14.	*Region*	Kategorical	Wilayah geografis pengguna (1, 2, 3, dll.)
15.	*TrafficType*	Kategorical	Jenis sumber lalu lintas (1, 2, 3, dll.)
16.	*VisitorType*	Kategorical	Jenis pengunjung : Returning_Visitor, New_Visitor, atau Other
17.	*Weekend*	Kategorical	Indikator apakah kunjungan dilakukan pada akhir pekan (True/False)
18.	*Revenue*	Kategorical	Target variabel, apakah terjadi pembelian (True/False)

Penelitian sebelumnya mengaplikasikan beberapa algoritma pembelajaran mesin seperti *Logistic Regression, Naive Bayes, Decision Tree, Random Forest*, dan *Gradient Boosting* untuk memprediksi niat pembelian online shopper berdasarkan data klikstream yang dikumpulkan melalui *Google Analytics*[6]. Penelitian menunjukkan bahwa *Decision Tree* memberikan nilai akurasi 81,8%.

Algoritma *LightGBM* sering digunakan untuk kasus prediksi dan klasifikasi. *LightGBM* digunakan untuk memprediksi intensitas curah hujan, menggunakan evaluasi matriks seperti akurasi, AUC, recall, precision, dan F1 score, dengan rata-rata akurasi 0,7251 [7]. Penelitian ini membandingkan metode *LightGBM* dan *XGBoost*, menunjukkan bahwa *XGBoost* unggul dalam akurasi dan sensitivitas, sementara *LightGBM* lebih baik dalam menebak kelas minoritas dengan rata-rata spesifisitas 80,41% dibandingkan 74,64% dari *XGBoost*[8]. Prediksi rasio klik tayang menggunakan *LightGBM* pada *dataset social network* menunjukkan hasil terbaik tanpa resampling dan dengan resampling ROS, dengan akurasi 91,25%, recall 93,10%, precision 84,38%, F1-Score 88,52%, dan AUC 0,92 [9].

Ketidakseimbangan kelas (*imbalanced class*) seringkali menjadi tantangan dalam pengolahan data, terutama dalam bidang klasifikasi. Ketidakseimbangan ini dapat menyebabkan algoritma pembelajaran mesin, seperti *LightGBM*, memberikan performa yang bias terhadap kelas mayoritas, sehingga mengurangi akurasi prediksi pada kelas minoritas. Untuk mengatasi masalah ini, teknik oversampling seperti SMOTE-ENN (Synthetic Minority Oversampling Technique - Edited Nearest Neighbors) dan Borderline SMOTE digunakan digunakan dalam penelitian ini. Pertanyaan penelitian sebagai berikut:

a) Bagaimana pengaruh implementasi SMOTE-ENN dan Borderline-SMOTE terhadap performa *LightGBM* pada dataset dengan kelas yang tidak seimbang?
b) Bagaimana performa Borderline SMOTE dibandingkan dengan SMOTE-ENN dalam meningkatkan kemampuan prediksi *LightGBM*?

METODE

Dataset

Data dalam penelitian ini menggunakan data sekunder yaitu dataset *online shoppers' intention* yang tersedia secara public di *UCI Machine Learning Repository*. Data terdiri dari 12.330 observasi dan 18 fitur yang dapat dikelompokan menjadi beberapa kategori seperti pada Tabel 1. Data tersebut digunakan untuk menganalisis perilaku pengguna pada situs *e-commerce*, khususnya untuk mengklasifikasikan apakah pengunjung akan melakukan pembelian (*Revenue = True*) atau tidak (*Revenue = False*). Selain itu, dataset ini sering digunakan untuk mempelajari perilaku konsumen berbasis data. Data mencakup pengamatan dari berbagai atribut yang relevan dengan keputusan pembelian.

Alur Penelitian

DOI : https://doi.org/10.36341/rabit.v10i1.5436 *Corresponding Author* : Yabes Dwi Nugroho H

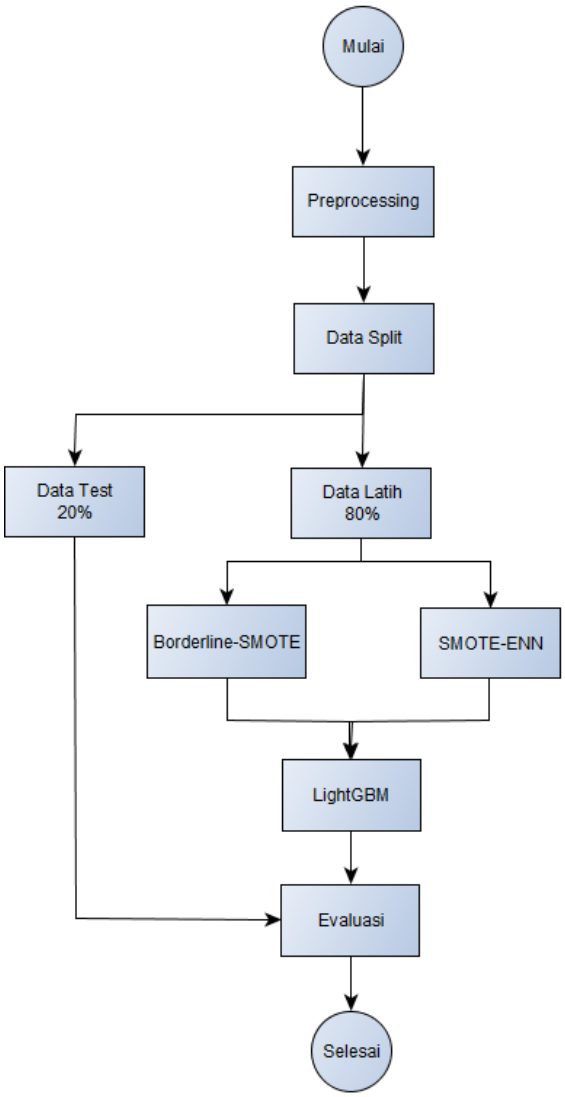
Gambar 1 menjelaskan alur penelitian untuk menangani ketidakseimbangan kelas dalam dataset. Proses dimulai dengan tahap *preprocessing*, Dimana data akan diproses untuk memastikan kualitasnya, seperti *data cleaning*, *encoding* dan standarisasi[10]. standarisasi data dilakukan untuk menghilangkan pengaruh skala yang berbeda-beda dari berbagai komponen pada ukuran langkah[11]. Standarisasi memastikan nilai rata-rata 0 dan deviasi standar 1, yang bermanfaat untuk analisis statistik[12]. Persamaan standardisasi yang digunakan dalam penelitian ini dapa dilihat sebagai berikut:

$$X_{new} = \frac{X-\mu}{\sigma} \qquad (1)$$

Dimana X adalah data, μ adalah rata-rata dan σ adalah standar deviasi. X_{new} merepresentasikan data standarisasi setelah transformasi.

Langkah selanjutnya, split data akan membagi data menjadi dua subset utama, yaitu 80 % untuk data latih dan 20 % untuk data uji. Pembagian ini bertujuan untuk menyediakan data pelatihan bagi model dan data yang terpisah untuk mengevaluasi performa model.Tahapan selanjutnya adalah penerapan *balancing data*. Data latih diproses lebih lanjut menggunakan teknik *resampling* seperti *SMOTE-ENN* dan *Borderline SMOTE*. Teknik-teknik ini bertujuan untuk menangani ketidakseimbangan kelas dengan menghasilkan sampel sintetis yang meningkatkan representasi kelas minoritas dan menghilangkan sampel yang redundan atau tidak relevan.

Data hasil resampling kemudian digunakan untuk melatih model *LightGBM*. Setelah pelatihan selesai, model diuji menggunakan *data uji*, dan hasil prediksi dievaluasi dengan metrik evaluasi yang sesuai, seperti akurasi, presisi, *recall*, atau *F1-score*. Proses ini memastikan bahwa model dapat menangani ketidakseimbangan data secara efektif sekaligus memberikan prediksi yang andal.



Gambar 1. Alur Penelitian

Borderline-SMOTE

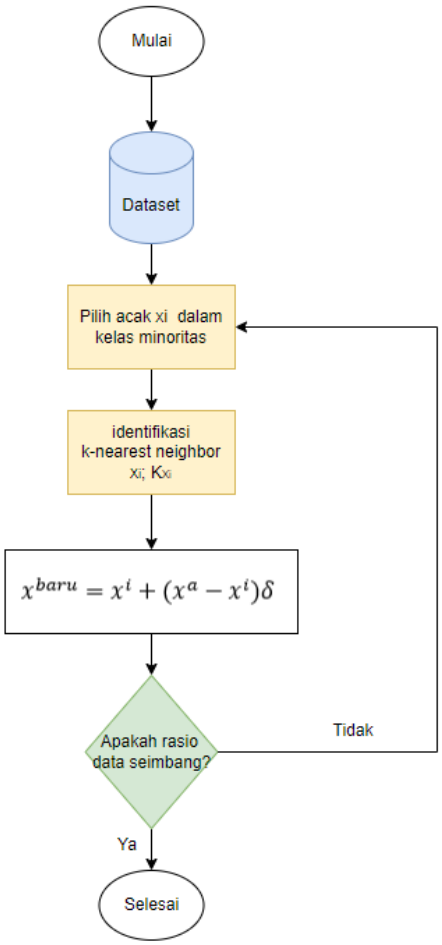
Borderline-SMOTE adalah salah satu pengembangan dari *Synthetic Minority Over-Sampling* (SMOTE) yang dirancang untuk menangani ketidakseimbangan kelas pada dataset dengan memberikan perhatian khusus pada sampel kelas minoritas yang berada didekat area *decision boundary*[13]. Borderline SMOTE bekerja dengan mengidentifikasi sampel minoritas yang beresiko tinggi terjadi kesalahan klasifikasi karena berada dekat dengan kelas mayoritas[14]. Data sintetis yang dihasilkan melalui interpolasi antara sampel minoritas terpilih dan tetangga minoritas

DOI : https://doi.org/10.36341/rabit.v10i1.5436 *Corresponding Author* : Yabes Dwi Nugroho H

terdekatnya, sehingga mencipatkan distribusi data yang lebih seimbang disekitar *decision boundary*[15].

SMOTE-ENN

Synthetic Minority Over-Sampling with Edited Nearest Neighbors (SMOTE-ENN) adalah sebuah algortima yang menggabungkan algoritma SMOTE dan ENN serta bekerja secara paralel untuk menangani masalah ketidakseimbangan kelas dalam dataset[16]. Metode ini mengkombinasikan teknik *oversampling* dan *undersampling*. Proses SMOTE dengan menggunakan *oversampling* memberikan sampel sintetis baru untuk kelas minoritas dengan melakukan interpolasi antara sampel minoritas yang ada[17].



Gambar 2. Proses SMOTE-ENN

Gambar 2 menunjukan proses ini SMOTE-ENN bertujuan untuk meningkatkan jumlah sampel kelas minoritas secara representatif tanpa menduplikasi data yang sudah ada. Setelah itu, metode ENN yang merupakan teknik *undersampling* akan menghapus sampel data yang dianggap sebagai *noise* atau tidak perlu atau ambiguitas berdasarkan evaluasi *k*-tetangga terdekat (*k-nearest neighbors*)[18][11]. Persamaan proses SMOTE-ENN ditunjukan sebagai berikut:

$$d(x^i, x^j) = \sqrt{\sum_{j=1}^{F}(x^i - x^j)^2} \qquad (2)$$

Dimana F adalah dimensi vektor fitur x^i, x^i sebagai pusat, pilih k sampel dengan jarak *ecludian* terdekat dan dapatkan k tetangga terdekat dari x^i [19].

Pilih secara acak sebuah sampel x^a dari k tetangga terdekat, hitung interpolasi linier antara x^i dan x^a, untuk mendapatkan sampel x^{baru} yang mendekati x^i, berikut persamaannya:

$$x^{baru} = x^i + (x^a - x^i)\delta \qquad (3)$$

Kombinasi ini menghasilkan dataset yang seimbang dan bersih, sehingga membantu model untuk mempelajari pola yang baik dari kedua kelas (mayoritas dan minoritas). Keunggulan *SMOTE-ENN* secara paralel mampu menyeimbangkan dataset dan meningkatkan kualitas data melalui penghapusan noise dan sampel yang salah. Sehingga, teknik ini tidak hanya mengatasi masalah ketidakseimbangan, namun meningkatkan akurasi dan generalisasi model.

LightGBM

Light Gradient Boosting Machine (*LightGBM*) merupakan algoritma pembelajaran mesin berbasis *gradient boosting* yang dirancang untuk meningkatkan efisiensi dan akurasi, terutama dalam menangani dataset berukuran besar dan berdimensi tinggi[20]. Algoritma ini menggunakan pendekatan inovatif seperti pembagian *leaf-wise* dalam pembuatan pohon keputusan, yang memungkinkan pengurangan kerugian lebih

DOI : https://doi.org/10.36341/rabit.v10i1.5436
Corresponding Author : Yabes Dwi Nugroho H

optimal dibandingkan metode pembagian *level-wise* tradisional. *LightGBM* juga mendukung berbagai teknik optimisasi, seperti *histogram-based decision tree* learning untuk mempercepat proses pelatihan dan mengurangi kebutuhan memori. Persamaannya sebagai berikut

$$F(x) = \sum_{m=1}^{M} f_m(x) \qquad (4)$$

Dimana $F(x)$ adalah hasil akhir, dan $f_m(x)$ adalah output dari pohon regresi ke-*m*[21]. Selain itu, algoritma ini mampu menangani fitur kategori secara efisien tanpa perlu dilakukan *encoding* manual. Keunggulan utama *LightGBM* meliputi kecepatan pelatihan yang tinggi, skalabilitas yang baik, dan kemampuan menangani ketidakseimbangan data atau *outlier*. Karakter yang ringan dan efisien dari *LightGBM* sering menjadi pilihan untuk berbagai aplikasi seperti prediksi keuangan, deteksi anomali, serta sistem rekomendasi. Meskipun demikian, performanya sangat bergantung pada tuning hyperparameter yang optimal untuk dataset tertentu.

Evaluasi

Evaluasi penelitian ini menggunakan confusion matrix. Confusion matrix adalah representasi tabel yang digunakan untuk mengevaluasi performa model klasifikasi dengan membandingkan prediksi model terhadap data sebenarnya[22]. Confusion matrix memberikan informasi yang sangat penting untuk menghitung metrik evaluasi seperti akurasi, presisi, *recall*, *F1-score*, dan *specificity*[23]. Persamaan evaluasi dijelaskan sebagai berikut:

$$Akurasi = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (5)$$

$$Presisi = \frac{TP}{(TP+FP)} \qquad (6)$$

$$Recall = \frac{TP}{(TP+FN)} \qquad (7)$$

$$F1 - Score = 2 \times \frac{presisi \times recall}{presisi+recall} \qquad (8)$$

Matriks ini terdiri dari empat elemen utama: *True Positiv*e (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

HASIL

Balancing Data

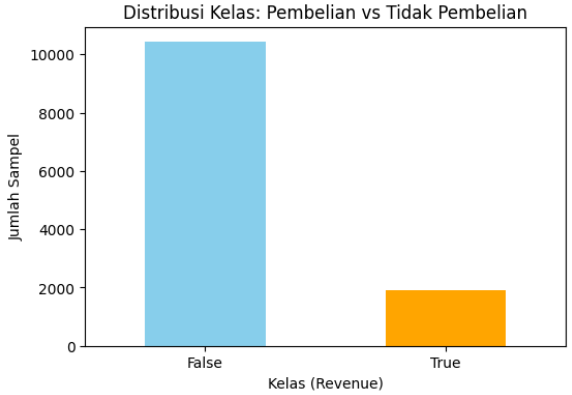
Gambar 3 menunjukkan distribusi kelas untuk variabel target, yaitu "pembelian" (*True*) versus "tidak pembelian" (*False*). Terlihat bahwa jumlah sampel untuk kelas "tidak pembelian" (*False*) sangat mendominasi jumlah sampel pada kelas "pembelian" (*True*). Distribusi ini menunjukkan adanya ketidakseimbangan data yang perlu ditangani untuk mencegah bias model terhadap kelas mayoritas.

Distribusi kelas sebelum dan sesudah penerapan metode SMOTE-ENN dapat dilihat pada Gambar 4. Sebelum dilakukan SMOTE-ENN, terdapat ketidakseimbangan data, di mana jumlah sampel pada kelas mayoritas (Kelas 0) jauh lebih besar dibandingkan kelas minoritas (kelas 1). Setelah SMOTE-ENN diterapkan, distribusi kelas menjadi lebih seimbang. Hal ini menunjukkan efektivitas metode SMOTE-ENN dalam menangani ketidakseimbangan data dengan menggabungkan metode oversampling dan penghapusan data redundan.

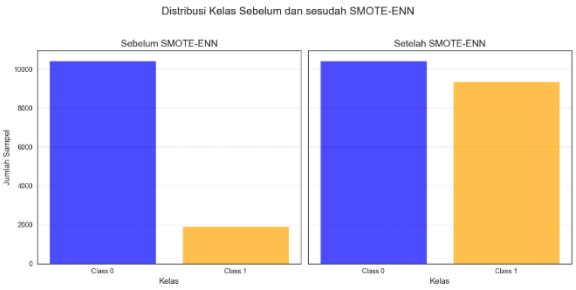
Distribusi kelas sebelum dan sesudah metode Borderline-SMOTE diterapkan dapat dilihat pada Gambar 5. Sebelum metode ini diterapkan, distribusi data menunjukkan dominasi kelas mayoritas (kelas 0) yang signifikan dibandingkan kelas minoritas (kelas 1). Setelah penerapan Borderline-SMOTE, distribusi data menjadi seimbang. Metode ini tidak hanya menambah sampel pada kelas minoritas tetapi juga menargetkan titik data di

DOI : https://doi.org/10.36341/rabit.v10i1.5436 *Corresponding Author* : Yabes Dwi Nugroho H

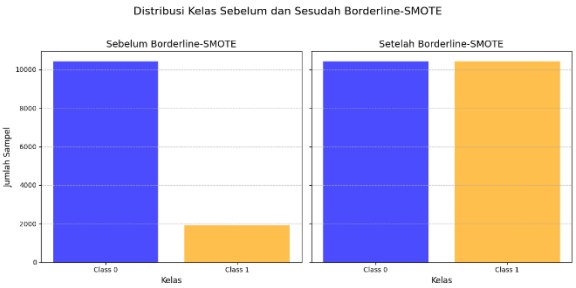
dekat batas keputusan untuk menghasilkan sampel sintetis yang lebih informatif.



Gambar 3. Ketidakseimbangan dataset



Gambar 4. Perbandingan Sebelum dan Sesudah SMOTE-ENN



Gambar 5. Perbandingan Sebelum dan Sesudah Borderline SMOTE

Tabel 2. Hasil Penelitian

Model	Presisi	Recall	Akurasi	F1-score
LightGBM	83%	77%	90%	80%
SMOTE-ENN +LightGBM	92%	93%	93%	93%
Borderline SMOTE+Light GBM	92%	92%	92%	92%

Tabel 2 menjelaskan perbandingan antara hasil *LightGBM* (tanpa penerapan SMOTE-ENN dan Borderline-SMOTE), SMOTE-ENN+*LightGBM* dan Borderline-SMOTE+*LightGBM*. Berdasarkan hasil evaluasi model seperti yang ditunjukan pada Tabel 2, pendekatan SMOTE-ENN dengan algoritma *LightGBM* menunjukkan performa terbaik dengan akurasi sebesar 93%, *precision*, *recall*, dan *F1-score* makro rata-rata sebesar 93%. Metode ini mampu memberikan keseimbangan yang baik antara kedua kelas (*False* dan *True*), masing-masing dengan *F1-score* sebesar 0.93. Hal ini menunjukkan bahwa teknik SMOTE-ENN efektif dalam menangani ketidakseimbangan data serta meningkatkan generalisasi model. Di sisi lain, model tanpa penerapan SMOTE menghasilkan akurasi sebesar 90%, dengan *precision* makro 83%, *recall* makro 77%, dan *F1-score* makro 80%. Namun, model ini cenderung bias terhadap kelas mayoritas *(False*), ditunjukkan dengan *F1-score* kelas True yang hanya mencapai 0.65, sehingga performa prediksi untuk kelas minoritas kurang optimal. Sementara itu, metode Borderline SMOTE dengan *LightGBM* memberikan hasil yang cukup kompetitif dengan akurasi sebesar 92%, serta *precision*, *recall*, dan *F1-score* makro rata-rata sebesar 92%. Model ini mampu menjaga keseimbangan performa antar kelas, dengan *F1-score* masing-masing kelas sebesar 0.92.

KESIMPULAN

Kesimpulan dari penelitian yang telah dilakukan adalah Secara keseluruhan, metode SMOTE-ENN dengan *LightGBM* memberikan hasil terbaik dalam meningkatkan akurasi dan keseimbangan performa antar kelas dibandingkan dengan metode lain, diikuti oleh *Borderline SMOTE* yang juga menghasilkan performa yang baik. Metode tanpa teknik penyeimbangan data memiliki performa yang

DOI : https://doi.org/10.36341/rabit.v10i1.5436 *Corresponding Author* : Yabes Dwi Nugroho H

paling rendah karena kurang mampu menangani ketidakseimbangan data.

Untuk penelitian selanjutnya, SMOTE-ENN dan Borderline-SMOTE dapat digunakan dalam kasus *Fraud detection* terkait data transaksi atau *marketing campaign* untuk prediksi pelanggan. Selain itu studi lanjutan, disarankan untuk mengeksplorasi kombinasi teknik penyeimbangan data lainnya atau menggunakan metode berbasis deep learning yang memiliki kemampuan lebih baik dalam menangani data yang kompleks dan tidak seimbang.

DAFTAR PUSTAKA

[1] S. Yoke Cheng, I. Alisa Hussain, K. Apparavu, and N. Trianna Rosli, 'Factors Influencing Online Shopping Intention Among Malaysians: A Quantitative-Based Study', *Electronic Journal of Business and Management*, vol. 7, pp. 2550–1380, 2019.

[2] K. Aidai and L. Zhi Chao, 'Analysis on the Influencing Factors of Consumers' Online Shopping Intention', *IOSR Journal of Business and Management (IOSR-JBM)*, vol. 22, pp. 27–35, 2020, doi: 10.9790/487X-2204022735.

[3] R. D. Nugraheni, 'Determinants And Classifications Of Online Shopping Consumers' Purchase Intention In Indonesia', *Jurnal Ekonomi Bisnis dan Kewirausahaan*, vol. 13, no. 1, p. 1, Apr. 2024, doi: 10.26418/jebik.v13i1.61399.

[4] A. Purnama, A. Maulana Yusup, A. Wibowo, and D. Susilawati, 'Uji Algoritma Random Forest Pada Dataset Online Shoppers Purchasing Intention', *Jurnal IKRA-ITH Informatika*, vol. 5, no. 1, 2021.

[5] S. Ahsain and M. Ait Kbir, 'Predicting the client's purchasing intention using Machine Learning models', in *E3S Web of Conferences*, EDP Sciences, May 2022. doi: 10.1051/e3sconf/202235101070.

[6] S. Yadav and M. T. Student, 'Prediction Of Online Shopper's Buying Intention Using Algorithms Of Pyspark Mllib', 2023. [Online]. Available: www.ijcspub.org

[7] D. Kurniawan, F. Oktaviansyah Hidayat, and A. Hari Saputra, 'Analisis Performa Model Lightgbm Dalam Prediksi Intensitas Hujan Wilayah Stasiun Meteorologi Kelas 1 Kualanamu Performance Analysis Of Lightgbm Model In Predicting Rain Intensity In The Kualanamu Class 1 Meteorological Station Area', *Jurnal Penelitian Sains dan Teknologi Indonesia*, vol. 3, no. 1, 2024.

[8] P. Septiana Rizky, R. Haiban Hirzi, U. Hidayaturrohman, U. Hamzanwadi Selong Jl TGKH Muhammad Zainuddin Abdul Madjid Pancor, and L. Timur, 'Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang', 2022. [Online]. Available: www.unipasby.ac.id

[9] K. Handayani, 'Penerapan Light Gradient Boosting Dalam Prediksi Rasio Klik Tayang', 2023.

[10] A. Widianti and I. Pratama, 'Penanganan Missing Values Dan Prediksi Data Timbunan Sampah Berbasis Machine Learning', *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 9, no. 2, pp. 242–251, Jul. 2024, doi: 10.36341/rabit.v9i2.4789.

[11] X. Jiao and J. Li, 'An Effective Intrusion Detection Model for Class-imbalanced Learning Based on SMOTE and Attention Mechanism', in *2021 18th International Conference on Privacy, Security and Trust, PST 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/PST52912.2021.9647756.

[12] N. M. Nayan, A. Islam, M. U. Islam, E. Ahmed, M. M. Hossain, and M. Z. Alam, 'SMOTE Oversampling and Near Miss Undersampling Based Diabetes Diagnosis from Imbalanced Dataset with XAI Visualization', in *Proceedings - IEEE Symposium on Computers and Communications*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ISCC58397.2023.10218281.

[13] J. Pardede and D. Prasetia Pamungkas, 'The Impact of Balanced Data Techniques on Classification Model Performance', *Scientific Journal of Informatics*, vol. 11, no. 2, 2024, doi: 10.15294/sji.v11i2.3649.

[14] N. Azmi Verdikha, T. Bharata Adji, A. Erna Permanasari, and D. Teknik Elektro dan Teknologi Informasi, 'Komparasi Metode Oversampling Untuk Klasifikasi Teks Ujaran Kebencian', *Seminar Nasional Teknologi Informasi dan Multimedia 2018,UNIVERSITAS AMIKOM Yogyakarta*, 2018.

[15] C. Jiang, W. Lv, and J. Li, 'Protein-Protein Interaction Sites Prediction Using Batch Normalization Based CNNs and Oversampling Method Borderline-SMOTE', *IEEE/ACM Trans Comput Biol Bioinform*, vol. 20, no. 3, pp. 2190–2199, May 2023, doi: 10.1109/TCBB.2023.3238001.

[16] H. Mohamed Ashif and E. G. M. Kanaga, 'Enhancing Diagnosis Precision in Alcohol Addiction Detection Through CNN Analysis with SMOTE-ENN Data Augmentation', in *Proceedings - 2024 4th International Conference on Pervasive Computing and Social Networking,*

ICPCSN 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 520–526. doi: 10.1109/ICPCSN62568.2024.00088.

[17] M. A. Latief, L. R. Nabila, W. Miftakhurrahman, S. Ma'rufatullah, and H. Tantyoko, 'Handling Imbalance Data using Hybrid Sampling SMOTE-ENN in Lung Cancer Classification', *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 3, no. 1, pp. 11–18, Feb. 2024, doi: 10.30812/ijecsa.v3i1.3758.

[18] I. Riantika, B. Sartono, and K. Anwar Notodiputro, 'Effectiveness of SMOTE-ENN to Reduce Complexity in Classification Model', *Indonesian Journal of Statistics and Its Applications*, vol. 8, no. 1, pp. 70–82, Jun. 2024, doi: 10.29244/ijsa.v8i1p70-82.

[19] T. Le, M. T. Vo, B. Vo, M. Y. Lee, and S. W. Baik, 'A Hybrid Approach Using Oversampling Technique and Cost-Sensitive Learning for Bankruptcy Prediction', *Complexity*, vol. 2019, 2019, doi: 10.1155/2019/8460934.

[20] T. O. Omotehinwa, D. O. Oyewola, and E. G. Moung, 'Optimizing the light gradient-boosting machine algorithm for an efficient early detection of coronary heart disease', *Informatics and Health*, vol. 1, no. 2, pp. 70–81, Sep. 2024, doi: 10.1016/j.infoh.2024.06.001.

[21] J. Ren, Z. Yu, G. Gao, G. Yu, and J. Yu, 'A CNN-LSTM-LightGBM based short-term wind power prediction method based on attention mechanism', *Energy Reports*, vol. 8, pp. 437–443, Aug. 2022, doi: 10.1016/j.egyr.2022.02.206.

[22] E. Novianto, A. Hermawan, and D. Avianto, 'Klasifikasi Algoritma K-Nearest Neighbor, Naive Bayes, Decision Tree Untuk Prediksi Status Kelulusan Mahasiswa S1', *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 8, no. 2, pp. 146–154, Jul. 2023, doi: 10.36341/rabit.v8i2.3434.

[23] A. Agustin, S. Andrean, S. Susanti, R. Rahmiati, and H. Hamdani, 'Review Aplikasi Kredivo Menggunakan Analisis Sentimen Dengan Algoritma Support Vector Machine', *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 9, no. 1, pp. 39–49, Dec. 2023, doi: 10.36341/rabit.v9i1.4107.