

Optimasi Bayesian pada Gradient Boosting untuk Prediksi Niat Beli E-Commerce pada Dataset dengan Ketidakseimbangan Kelas

Imam Bagus Setyawan*, Heribertus Himawan

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ^{1,*}111202012526@mhs.dinus.ac.id, ²himawan26@dsn.dinus.ac.id

Email Penulis Korespondensi: 111202012526@mhs.dinus.ac.id

Submitted: 20/04/2026; Accepted: 02/06/2026; Published: 05/06/2026

Abstrak—Prediksi niat beli (purchase intention) konsumen dalam e-commerce merupakan tantangan krusial akibat tingginya tingkat ketidakseimbangan kelas (class imbalance), di mana mayoritas pengunjung hanya melakukan penelusuran tanpa transaksi. Penelitian ini mengomparasikan kinerja tiga algoritma keluarga Gradient Boosting (XGBoost, LightGBM, dan CatBoost) menggunakan dataset Online Shoppers Intention dengan rasio kelas 84,5% berbanding 15,5%. Untuk mengatasi bias kelas mayoritas, metode Synthetic Minority Oversampling Technique (SMOTE) diimplementasikan pada data latih. Penelitian ini berfokus pada implementasi optimasi hyperparameter menggunakan framework Optuna berbasis Tree-structured Parzen Estimator (TPE), yang divalidasi secara statistik melalui uji Friedman dan Post-Hoc Nemenyi. Evaluasi model menggunakan stratified 10-Fold Cross-Validation menunjukkan bahwa ketiga model mampu menangani ketidakseimbangan kelas secara efektif. LightGBM mencapai akurasi 88,36% dengan ROC-AUC 0,9138, XGBoost mencapai akurasi 88,56% dengan ROC-AUC 0,9127, dan CatBoost mencapai akurasi 88,56% dengan ROC-AUC 0,9121. Analisis feature importance mengidentifikasi ProductRelated_Duration dan ExitRates sebagai prediktor utama niat beli. Uji statistik Friedman mendeteksi adanya perbedaan kinerja secara global ($p=0,0450$), namun uji lanjutan Nemenyi tidak menemukan bukti empiris yang cukup untuk menyatakan perbedaan performa yang signifikan secara berpasangan. Penelitian ini memberikan kontribusi praktis bagi industri e-commerce dengan mendemonstrasikan bahwa pemilihan algoritma ansambel tidak lagi harus bergantung mutlak pada margin akurasi yang semu, melainkan dapat direkomendasikan secara objektif berdasarkan efisiensi latensi komputasi, di mana arsitektur LightGBM terbukti lebih efisien.

Kata Kunci: CatBoost; Ketidakseimbangan Kelas; LightGBM; Niat Beli; Optuna; XGBoost

Abstract—Predicting consumer purchase intention in e-commerce is a crucial challenge due to the high rate of class imbalance, where the majority of visitors only browse without making a transaction. This study compares the performance of three Gradient Boosting family algorithms (XGBoost, LightGBM, and CatBoost) using the Online Shoppers Intention dataset, which has a class ratio of 84.5% to 15.5%. To overcome majority class bias, the Synthetic Minority Oversampling Technique (SMOTE) approach was implemented on the training data. This research focuses on hyperparameter optimization implementation using the Optuna framework based on the Tree-structured Parzen Estimator (TPE), which is statistically validated using the Friedman and Post-Hoc Nemenyi tests. Model evaluation using stratified 10-Fold Cross-Validation shows that all three models can handle class imbalance effectively. LightGBM achieved an accuracy of 88.36% with an ROC-AUC of 0.9138, XGBoost achieved an accuracy of 88.56% with an ROC-AUC of 0.9127, and CatBoost achieved an accuracy of 88.56% with an ROC-AUC of 0.9121. Feature importance analysis identifies ProductRelated_Duration and ExitRates as the main predictors of purchase intention. The Friedman statistical test detected global performance differences ($p=0.0450$), but the Nemenyi post-hoc test found insufficient empirical evidence to claim significant pairwise performance differences. This research provides a practical contribution to the e-commerce industry by demonstrating that the selection of ensemble algorithms no longer needs to rely absolutely on pseudo-accuracy margins, but can be objectively recommended based on computational latency efficiency, where the LightGBM architecture proves to be efficient.

Keywords: CatBoost; Class Imbalance; LightGBM; Purchase Intention; Optuna; XGBoost

1. PENDAHULUAN

Perkembangan pesat ekosistem perdagangan elektronik (e-commerce) dalam satu dekade terakhir telah mengubah paradigma analitik data secara fundamental, beralih dari pengolahan data yang sekadar bersifat deskriptif menuju implementasi analitik prediktif secara real-time. Di tengah kompetisi pasar digital yang semakin ketat, salah satu tantangan paling krusial bagi praktisi rekayasa data dan pelaku industri ritel bukanlah lagi pada pengukuran kepuasan pelanggan pasca-pembelian, melainkan pada kemampuan untuk memprediksi niat beli (purchase intention) konsumen sesaat sebelum pengunjung tersebut meninggalkan sesi navigasi mereka. Mayoritas literatur analitik e-commerce kontemporer saat ini masih cenderung menitikberatkan perhatian pada pendekatan Natural Language Processing (NLP) untuk mengekstraksi sentimen dari teks ulasan atau komentar pada toko daring. Meskipun analisis sentimen memberikan nilai strategis, pendekatan pasca-transaksi semacam ini secara inheren tidak memiliki kapabilitas untuk mencegah fenomena cart abandonment atau pengabaian keranjang belanja yang sering terjadi di tengah proses penelusuran. Oleh karena itu, memodelkan intensi konsumen dengan memanfaatkan data historis perilaku navigasi (clickstream) secara dinamis menyajikan tingkat kompleksitas yang jauh lebih tinggi. Data clickstream merekam jejak interaksi mikro seperti pergerakan kursor, durasi waktu yang dihabiskan pada kategori halaman tertentu, hingga tingkat pentalan (bounce rate) yang secara kolektif merepresentasikan kondisi kognitif dan tingkat keterlibatan pengunjung secara langsung [1].

Tantangan fundamental yang menyertai pemrosesan data clickstream dalam domain e-commerce adalah keberadaan fenomena ketidakseimbangan kelas (class imbalance) yang sangat ekstrem. Dalam skenario operasional dunia nyata, proporsi pengunjung yang hanya sekadar melakukan penelusuran katalog web tanpa diakhiri dengan transaksi final selalu mendominasi populasi data secara absolut. Pada dataset standar Online Shoppers Intention yang

secara konsisten diadopsi sebagai benchmark global dalam berbagai literatur, tercatat secara empiris bahwa sekitar 84,5% dari total pengunjung merepresentasikan kelas non-transaksi, sementara hanya 15,5% sisanya yang berujung pada penyelesaian transaksi[2]. Ketimpangan distribusi kelas yang ekstrem ini menjadi kendala teknis bagi algoritma machine learning konvensional, di mana model sering kali mengalami bias yang parah dengan cenderung memprediksi kelas mayoritas secara konsisten demi mencapai skor akurasi yang tinggi secara artifisial. Sebagai konsekuensinya, model tersebut akan gagal total dalam mendeteksi keberadaan kelas minoritas yang sejatinya merupakan target konsumen aktual bagi strategi pemasaran. Dalam konteks ini, implementasi teknik sintesis data seperti Synthetic Minority Oversampling Technique (SMOTE) bukan lagi sekadar pilihan tambahan, melainkan telah menjadi standar intervensi metodologis yang wajib dilakukan untuk menyeimbangkan representasi kelas sebelum data dieksekusi oleh arsitektur model prediksi [3], [4].

Sejumlah penelitian terdahulu telah memberikan kontribusi signifikan dalam upaya menangani masalah prediksi niat beli ini melalui pemanfaatan berbagai variasi algoritma klasifikasi. Studi yang dipelopori oleh Wen et al. (2023) berhasil mendemonstrasikan efektivitas pendekatan berbasis machine learning untuk mengidentifikasi niat beli pada konsumen anonim melalui data clickstream, sekaligus menegaskan bahwa variabel log navigasi memiliki bobot prediktif yang jauh lebih signifikan dibandingkan data demografis yang bersifat statis [1]. Selaras dengan temuan tersebut, Mishra et al. (2025) mengajukan arsitektur imbalance-aware pipelines yang dirancang khusus untuk sektor e-commerce, dengan hasil yang mengonfirmasi bahwa penanganan ketidakseimbangan kelas memegang peranan vital dalam menekan angka kesalahan prediksi pada fase kritis keranjang belanja [5]. Ruscikarani et al. (2025) kemudian memperluas cakupan penelitian dengan mengeksplorasi penggunaan algoritma ansambel mutakhir Extreme Gradient Boosting (XGBoost) untuk memprediksi intensi berdasarkan perilaku sesi pengguna secara mendalam [6]. Di sisi lain, sebuah riset komparatif terbaru yang dipublikasikan oleh Ajibade et al. (2025) mengimplementasikan SMOTE secara beriringan dengan model Logistic Regression, XGBoost, dan Random Forest pada dataset yang sama. Menariknya, studi tersebut menyimpulkan bahwa algoritma Random Forest konvensional mampu mengungguli varian XGBoost dalam menangkap korelasi non-linear pada perilaku pengunjung e-commerce [7].

Meskipun fondasi teoritis dari berbagai studi tersebut telah terbangun dengan kuat, penelitian ini mengidentifikasi adanya celah penelitian (research gap) yang mendasar terkait dengan metodologi evaluasi algoritma yang tidak proporsional. Kesimpulan mengenai superioritas satu algoritma atas algoritma lainnya—seperti klaim yang menempatkan Random Forest di atas algoritma keluarga boosting—sering kali rentan terhadap bias evaluasi (biased comparison). Hal ini umumnya terjadi karena ruang pencarian hiperparameter (hyperparameter search space) dari model-model boosting yang kompleks sering kali tidak dioptimasi secara maksimal hingga mencapai batas performa puncaknya [8]. Sebagian besar literatur pendahulu masih sangat bergantung pada penggunaan parameter bawaan (default) atau mengandalkan teknik Grid Search klasik yang memiliki keterbatasan jangkauan ruang pencarian dan sangat mahal dari segi waktu komputasi. Sampai saat ini, belum ditemukan kajian saintifik yang secara berkeadilan membandingkan performa puncak dari tiga generasi mutakhir dalam keluarga Gradient Boosting—yakni XGBoost, LightGBM, dan CatBoost [9] pada kondisi ketidakseimbangan kelas ekstrem yang dikawinkan dengan kerangka optimasi berbasis Bayesian [10], [11].

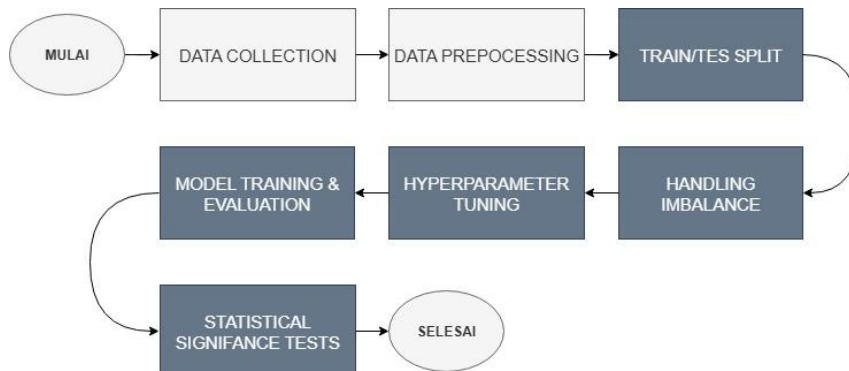
Sebagai respons terhadap kesenjangan literatur tersebut, penelitian ini tidak dirancang untuk mengajukan sebuah desain algoritma klasifikasi baru, melainkan untuk memperkenalkan sebuah metodologi evaluasi komparatif yang objektif, transparan, dan teroptimasi secara penuh. Proses penyyetelan hiperparameter dalam riset ini dieksekusi secara sistematis menggunakan framework Optuna yang berbasis pada algoritma Tree-structured Parzen Estimator (TPE) [12]. Pendekatan ini dipilih guna memastikan bahwa setiap varian algoritma boosting yang diuji benar-benar telah didorong hingga batas kapasitas komputasi puncaknya sebelum dibandingkan kinerjanya secara statistik. Melalui integrasi pipeline yang tangguh antara metode SMOTE dan keluarga Gradient Boosting, penelitian ini berkontribusi dalam menetralkan bias kelas mayoritas secara efektif pada data log clickstream yang bersifat volatil dan tidak pasti. Selain itu, penelitian ini juga memberikan pembaruan dalam aspek keadilan komparasi berbasis Bayesian yang mampu mengeliminasi potensi perbandingan performa yang bersifat bias akibat optimasi parameter yang tidak memadai. Hasil akhir dari penelitian ini diharapkan dapat memberikan landasan validasi objektif bagi praktisi industri melalui penyajian rekomendasi strategis yang didasarkan pada analisis statistik inferensial, seperti uji Friedman dan uji post-hoc Nemenyi [13]. Fokus utama dari rekomendasi tersebut tidak hanya berpaku pada margin tipis angka akurasi persentase semata, namun juga memprioritaskan aspek efisiensi sumber daya komputasi yang sangat krusial bagi implementasi sistem prediksi di lingkungan produksi e-commerce skala besar.

2. METODOLOGI PENELITIAN

Metodologi pada penelitian ini dirancang menggunakan pendekatan kuantitatif eksperimental yang mencakup tahapan pemrosesan data ujung-ke-ujung (*end-to-end*). Tujuan dari metodologi ini adalah untuk memastikan bahwa penanganan ketidakseimbangan kelas (*class imbalance*) pada data historis clickstream e-commerce dapat diselesaikan secara optimal sebelum dieksekusi oleh algoritma machine learning. Proses ini dirancang secara terstruktur mulai dari akuisisi data, prapemrosesan, penanganan data yang tidak seimbang menggunakan teknik oversampling, penyyetelan hiperparameter (hyperparameter tuning), hingga validasi secara statistik [3].

2.1 Tahapan Penelitian

Tahapan penelitian merupakan langkah-langkah sistematis yang dilalui dalam proses pengembangan model prediksi niat beli konsumen. Gambar 1 mengilustrasikan alur kerja (*flowchart*) penelitian secara keseluruhan. Report, Confusion Matrix, dan kurva ROC-AUC, yang kemudian diakhiri dengan Uji Signifikansi Statistik (Uji Friedman dan Post-Hoc Nemenyi) untuk menarik kesimpulan yang valid secara saintifik.



Gambar 1. Tahapan penelitian

Berdasarkan Gambar 1, penjelasan untuk masing-masing tahapan dalam penelitian ini adalah sebagai berikut:

- Pengumpulan Data (Data Collection): Mengakuisisi dataset Online Shoppers Intention [2] yang memuat log perilaku pengunjung e-commerce.
- Pra-pemrosesan Data (Data Preprocessing): Melakukan pembersihan data, konversi tipe data boolean ke integer, dan menerapkan Label Encoding pada variabel kategorikal tekstual seperti jenis pengunjung dan bulan.
- Pembagian Data (Train/Test Split): Memisahkan dataset menjadi set data latih (80%) dan set data uji (20%) menggunakan metode stratified sampling agar rasio kelas minoritas dan mayoritas tetap proporsional.
- Penanganan Data Tidak Seimbang (Handling Imbalance): Menerapkan metode Synthetic Minority Oversampling Technique (SMOTE) secara eksklusif hanya pada set data latih untuk mensintesis sampel kelas minoritas sehingga rasio kelas menjadi seimbang (1:1).
- Optimasi Hiperparameter (Hyperparameter Tuning): Menggunakan framework Optuna dengan algoritma Tree-structured Parzen Estimator (TPE) untuk mencari kombinasi parameter terbaik dari masing-masing algoritma klasifikasi [10].
- Pelatihan dan Evaluasi Model (Model Training & Evaluation): Melatih model XGBoost, LightGBM, dan CatBoost menggunakan parameter optimal, kemudian mengevaluasinya dengan metrik klasifikasi seperti Accuracy, Precision, Recall, F1-Score, dan ROC-AUC [14], [15].
- Uji Signifikansi Statistik: Melakukan validasi silang (10-Fold CV) dan mengujinya menggunakan uji statistik Friedman serta uji Post-Hoc Nemenyi untuk membuktikan signifikansi perbedaan performa antar algoritma secara saintifik [13], [16], [17].

Adapun sumber data yang digunakan pada tahapan awal pengumpulan data adalah dataset sekunder. Penelitian ini menggunakan dataset sekunder *Online Shoppers Intention* [2], yang merekam 12.330 sesi interaksi pengguna dalam sebuah platform *e-commerce*. Dataset ini berdimensi 18 kolom, yang terdiri dari fitur numerik dan kategorikal. Karakteristik lengkap dan deskripsi fitur dari dataset yang digunakan disajikan pada Tabel 1.

Tabel 1. Karakteristik dan Deskripsi Fitur Dataset

Nama Fitur	Tipe Data	Deskripsi
Administrative	Integer	Jumlah halaman bertipe administratif (seperti akun pengguna atau manajemen profil) yang dikunjungi oleh pengunjung dalam satu sesi.
Administrative_Duration	Float	Total durasi waktu (dalam detik) yang dihabiskan pengunjung pada kategori halaman administratif selama sesi berlangsung.
Informational	Integer	Akumulasi jumlah halaman bertipe informasi (seperti "Tentang Kami", "Kontak", atau "Alamat Toko") yang diakses pengunjung
Informational_Duration	Float	Total durasi waktu yang dihabiskan pengunjung untuk membaca atau berinteraksi pada kategori halaman informasi
ProductRelated	Integer	Jumlah halaman terkait produk (katalog, detail produk, atau spesifikasi) yang dilihat pengunjung; sering menjadi indikator utama minat beli.
ProductRelated_Duration	Float	Total waktu yang dihabiskan untuk menelusuri produk; fitur ini teridentifikasi sebagai prediktor paling dominan dalam model.
BounceRates	Float	Rata-rata persentase pengunjung yang masuk ke halaman tertentu dan langsung keluar tanpa melakukan aksi atau interaksi lebih lanjut.



Nama Fitur	Tipe Data	Deskripsi
ExitRates	Float	Persentase jumlah kunjungan yang berakhir pada halaman tersebut dibandingkan dengan total tayangan halaman (pageviews) dalam satu sesi
PageValues	Float	Nilai rata-rata halaman web yang dikunjungi pengguna sebelum menyelesaikan transaksi e-commerce; memiliki korelasi positif kuat terhadap Revenue.
SpecialDay	Float	Nilai kedekatan tanggal kunjungan dengan hari spesial (misalnya Hari Valentine atau Lebaran). Nilai mendekati 1 menunjukkan sesi terjadi sangat dekat dengan hari tersebut.
Month	Categorical	Nama bulan saat sesi kunjungan terjadi (Januari hingga Desember).
OperatingSystems	Categorical	Kode identifikasi sistem operasi yang digunakan oleh pengunjung saat mengakses situs.
Browser	Categorical	Kode identifikasi jenis peramban (<i>browser</i>) yang digunakan pengunjung.
Region	Categorical	Kode wilayah geografis asal pengunjung berdasarkan alamat IP atau data profil.
TrafficType	Categorical	Kode jenis trafik yang membawa pengunjung ke situs (misalnya melalui iklan, pencarian organik, atau media sosial).
VisitorType	Categorical	Klasifikasi tipe pengunjung, yang dibedakan menjadi Pengunjung Baru (<i>New Visitor</i>), Pengunjung Lama (<i>Returning Visitor</i>), dan lainnya.
Weekend	Boolean	Indikator apakah sesi kunjungan terjadi pada akhir pekan (Sabtu/Minggu). Diolah menjadi biner (0 atau 1) saat prapemrosesan.
Revenue	Boolean	Variabel target yang menunjukkan apakah sesi tersebut berakhir dengan transaksi (True) atau tidak (False).

Dari 18 fitur di atas, 10 bersifat numerik dan 6 kategorikal, dengan Revenue sebagai variabel target biner yang akan diprediksi.

2.2 Pra-pemrosesan Data dan Implementasi SMOTE

Pada tahap prapemrosesan, fitur dengan tipe data boolean dikonversi menjadi format integer (0 dan 1). Variabel bertipe kategorikal (Month dan VisitorType) ditransformasikan ke dalam representasi numerik menggunakan metode Label Encoding. Dataset kemudian dipartisi menjadi variabel prediktor (X) berisi 17 fitur dan variabel target (y) yaitu Revenue. Pembagian data latih dan data uji dikonfigurasi dengan teknik Stratified Train-Test Split (80:20) untuk menjamin bahwa rasio ketidakseimbangan kelas awal tetap dipertahankan pada kedua partisi, menghasilkan 9.864 sampel latih (8.338 kelas False, 1.526 kelas True) dan 2.466 sampel uji (2.084 kelas False, 382 kelas True) [3].

Untuk mengatasi degradasi kinerja prediksi pada kelas minoritas akibat fenomena class imbalance, penelitian ini mengimplementasikan algoritma penyeimbangan data Synthetic Minority Oversampling Technique (SMOTE). SMOTE beroperasi dengan cara mensintesis data buatan (artificial) untuk kelas minoritas berdasarkan kedekatan jarak ruang fitur (Euclidean distance) menggunakan prinsip k-Nearest Neighbors [18] Formula matematis dasar penciptaan vektor sampel sintetis baru (x_{baru}) dirumuskan pada Persamaan (1):

$$x_{baru} = x_i + \lambda \times (x_{zi} - x_i) \tag{1}$$

Dalam penelitian ini, parameter $k_neighbors$ ditetapkan sebesar 5 ($k=5$). Pemilihan nilai ini didasarkan pada standar heuristik yang stabil untuk menjaga keseimbangan antara informasi lokal dan struktur global data, sehingga mencegah overfitting pada lingkungan tetangga yang terlalu sempit. Pembagian data dilakukan dengan teknik Stratified Train-Test Split (80:20) untuk menjamin rasio kelas awal tetap dipertahankan pada kedua partisi.

2.3 Algoritma Gradient Boosting dan Optimasi Optuna

Penelitian ini membandingkan tiga algoritma state-of-the-art: XGBoost, LightGBM, dan CatBoost. Untuk memaksimalkan kapabilitas masing-masing model, dilakukan hyperparameter tuning menggunakan framework Optuna dengan algoritma Tree-structured Parzen Estimator (TPE) [19], [20], [21]. Pemanfaatan TPE dipilih karena kemampuannya dalam memodelkan kepadatan probabilitas parameter secara Bayesian, yang terbukti jauh lebih efisien dalam mengeksplorasi ruang pencarian yang luas dibandingkan metode Grid Search konvensional [8], [22]. Ruang pencarian (search space) yang didefinisikan untuk proses optimasi disajikan pada Tabel 2.

Tabel 2. Konfigurasi Parameter Optimal Hasil Optuna (TPE)

Algoritma	Parameter Utama	Ruang Pencarian
LightGBM	Learning_rate, max_depth	0.01 – 0.3(log), 3 -12
	Num_leaves, min_child_samples	20 -150,5 - 100
CatBoost	Iterations, depth	100 -500,4 – 10

Algoritma	Parameter Utama	Ruang Pencarian
XGBoost	Learning_rate, 12_leaf_reg	0.01 – 0.3(log), 10 ⁻³ – 10(log)
	Learning_rate, max_depth	0.01 – 0.3(log), 3 - 10
	Subsample, colsample_bytree	0.5 – 1.0, 0.5 – 1.0

Konfigurasi tersebut merupakan representasi kombinasi nilai yang berhasil memberikan nilai validasi ROC-AUC rata-rata silang terbaik selama fase pencarian iteratif.

2.4 Metrik Evaluasi dan Uji Signifikansi Statistik

Kinerja model dievaluasi secara komprehensif menggunakan matriks kebingungan (Confusion Matrix) yang mengukur komponen True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN). Dari komponen tersebut, diturunkan persamaan metrik evaluasi klasifikasi sebagai berikut [27]:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{2}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{3}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{4}$$

$$F1 - Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{5}$$

Evaluasi kinerja model bertumpu pada metrik ROC-AUC yang tangguh terhadap ketidakseimbangan kelas, divalidasi menggunakan *10-Fold Stratified Cross-Validation* guna menjaga stabilitas sebaran sampel [14]. Signifikansi perbedaan performa kemudian dianalisis melalui uji non-parametrik Friedman dan *Post-Hoc* Nemenyi. Selain metrik diskriminatif tersebut, penelitian ini turut mengukur efisiensi waktu komputasi (*training latency*) sebagai parameter penentu kelayakan implementasi model pada skala industri [23].

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penanganan Ketidakseimbangan Data (SMOTE)

Sebelum model dilatih, analisis terhadap distribusi kelas target pada data latih dilakukan untuk memastikan kesiapan data. Seperti yang ditunjukkan pada Tabel 3, distribusi awal kelas target (*Revenue*) sangat timpang, di mana mayoritas pengunjung tidak melakukan transaksi.

Tabel 3. Distribusi Kelas Target (Revenue Awal)

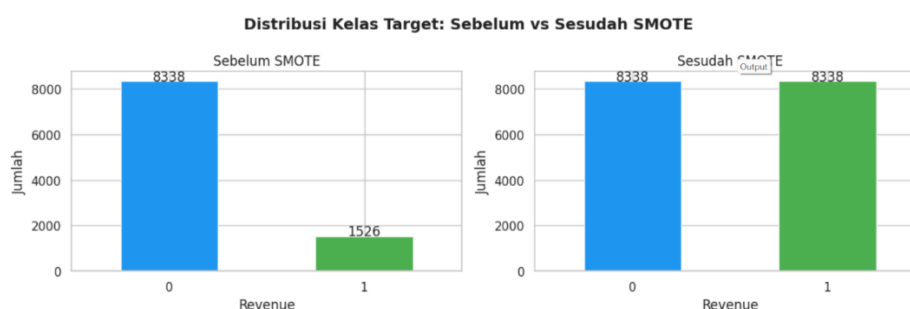
Kelas Target	Jumlah Pengunjung	Persentase
False (tidak beli)	10.422	84,5%
True (beli)	1.908	15,5%

Untuk mengatasi bias yang diakibatkan oleh dominasi kelas False, teknik SMOTE diterapkan secara eksklusif pada data latih. Tabel 4 menyajikan perbandingan detail jumlah sampel sebelum dan sesudah intervensi algoritma SMOTE dilakukan.

Tabel 4. Distribusi Kelas Data Latih Sebelum dan Sesudah SMOTE

Kelas Target	Sebelum SMOTE	Sesudah SMOTE
False (tidak beli)	8.338	8.338
True (beli)	1.908	8.338
Total	9.864	16.676

Secara visual, signifikansi transformasi penyeimbangan data latih ini dapat diamati lebih jelas pada Gambar 2.



Gambar 2 . Distribusi Kelas Target Data Latih: Sebelum vs Sesudah SMOTE

Gambar 2 tersebut mengilustrasikan bagaimana kelas minoritas disintesis secara proporsional hingga mencapai keseimbangan sempurna (8.338 sampel) atau rasio 1:1. Penyeimbangan ini secara analitis sangat krusial agar arsitektur ansambel Boosting tidak terjebak pada ilusi akurasi semu akibat memprediksi kelas mayoritas secara membabi buta pada tahap pelatihan.

3.2 Hasil Optimasi Hyperparameter (Optuna)

Proses pencarian hiperparameter menggunakan algoritma Tree-structured Parzen Estimator (TPE) dalam framework Optuna dieksekusi sebanyak 30 iterasi percobaan untuk masing-masing model. Selama proses iterasi, algoritma Bayesian secara cerdas mengarahkan evaluasi pada area ruang pencarian yang paling menjanjikan guna memaksimalkan fungsi objektif, yakni skor rata-rata Cross-Validation ROC-AUC.

Berdasarkan hasil eksekusi, Optuna berhasil mencapai konvergensi dengan sangat baik dan menemukan kombinasi hiperparameter paling optimal (Best Parameters) yang menyeimbangkan antara kompleksitas pohon (tree depth/leaves) dan laju pembelajaran (learning rate). Rincian hiperparameter terbaik yang dihasilkan oleh Optuna dan kemudian disuntikkan ke dalam model final disajikan pada Tabel 5.

Tabel 5. Nilai Hiperparameter Optimal Hasil Pencarian Optuna

Algoritma	Parameter	Nilai Optimal
LightGBM	learning_rate	0.21
	max_depth	12
	num_leaves	98
	min_child_samples	77
	feature_fraction	0.6812
	bagging_fraction	0.8672
CatBoost	learning_rate	0.133
	depth	9
	iterations	356
	l2_leaf_reg	0.4235
XGBoost	border_count	32
	learning_rate	0.1976
	max_depth	9
	subsample	0.9402
	colsample_bytree	0.8936
	min_child_weight	1

Berdasarkan Tabel 5, hasil pencarian ruang hiperparameter menunjukkan pola adaptasi yang unik dari masing-masing algoritma terhadap struktur dataset e-commerce. Pada algoritma LightGBM, kombinasi learning rate yang tergolong cukup adaptif (0,2100) diseimbangkan dengan pembentukan pohon yang dalam (max depth = 12) dan jumlah daun yang besar (num leaves = 98). Konfigurasi ini mengindikasikan bahwa model diizinkan untuk secara agresif mengekstraksi pola kompleks dari data log pengunjung, namun tetap dikunci dari risiko overfitting melalui pembatasan acak pada kolom fitur (feature fraction = 0,6812).

Sementara itu, CatBoost menemukan titik konvergensi optimalnya pada iterasi ke-356 dengan learning rate moderat (0,1330) dan sangat mengandalkan nilai regularisasi penalti (l2_leaf_reg = 0,4235) untuk meredam noise. Di sisi lain, XGBoost memaksimalkan kapabilitasnya dengan rasio subsample yang sangat tinggi (0,9402), yang berarti algoritma ini menggunakan hingga 94% porsi data latih pada setiap iterasi pembentukan pohonnya untuk memperkuat deteksi kelas minoritas. Secara keseluruhan, perolehan nilai-nilai hiperparameter ini mengeliminasi bias intervensi manual dari peneliti, sehingga menjamin bahwa ketiga model benar-benar beroperasi pada puncak kapasitas tertingginya sebelum saling dibandingkan pada tahapan selanjutnya.

3.3 Evaluasi Kinerja Klasifikasi Final

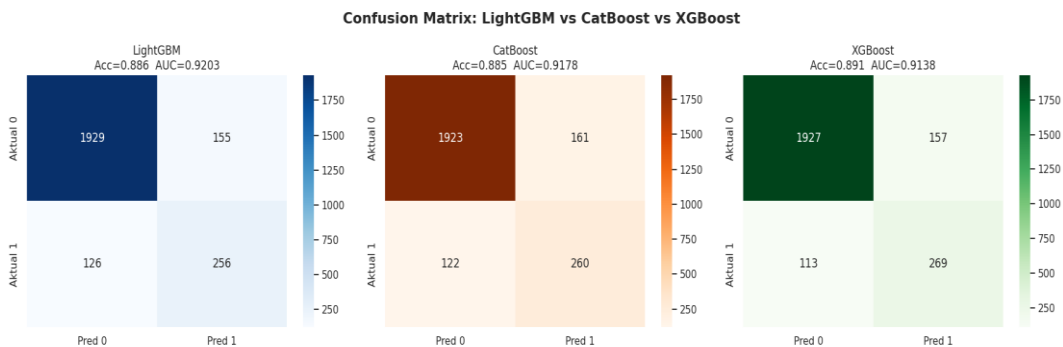
Model ansambel yang telah disuntikkan hiperparameter optimal divalidasi kinerjanya menggunakan 2.466 sampel data pengujian yang sama sekali belum pernah “dilihat” pada fase pelatihan. Tabel 6 menyajikan hasil Classification Report lengkap dari masing-masing algoritma beserta nilai ROC-AUC pada data uji.

Tabel 6. Hasil Classification Report pada Data Uji (n=2.466)

Model	Kelas	Precision	Recall	F1-Score	Accuracy	ROC-AUC	TP / TN / FP / FN
LightGBM	0 (tidak beli)	0,94	0,92	0,93	88,36%	0,9138	TN=1918 / TP=261 / FP=166 / FN=121
	1 (beli)	0,61	0,68	0,65	-	-	-
CatBoost	0 (tidak beli)	0,94	0,92	0,93	88,56%	0,9121	TN=1919 / TP=265 / FP=165 / FN=117
	1 (beli)	0,62	0,69	0,65	-	-	-

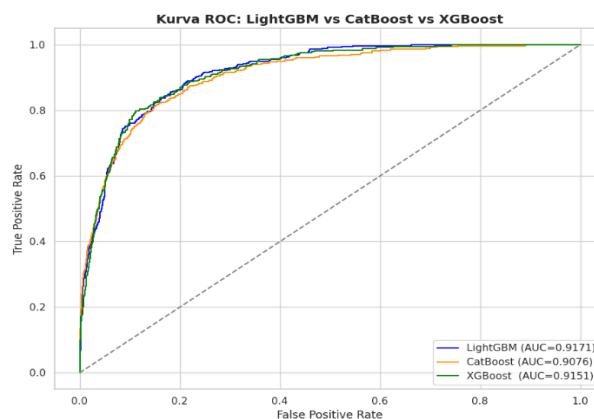
Model	Kelas	Precision	Recall	F1-Score	Accuracy	ROC-AUC	TP / TN / FP / FN
XGBoost	0 (tidak beli)	0,94	0,92	0,93	88,56%	0,9127	TN=1923 / TP=261 / FP=161 / FN=121
	1 (beli)	0,62	0,68	0,65	-	-	-

Berdasarkan Tabel 6, kinerja kemampuan diskriminatif model terlihat sangat stabil secara komparatif. Dari Confusion Matrix (Gambar 3), LightGBM mencatatkan True Positives sebanyak 261 sampel, CatBoost sebanyak 265 sampel (tertinggi), dan XGBoost sebanyak 261 sampel. Angka F1-Score yang tertahan di level 0,65 untuk kelas minoritas (True) merupakan cerminan langsung dari kesulitan bawaan dalam memisahkan intensi pengunjung yang hanya memasukkan produk ke keranjang namun gagal bertransaksi, dengan intensi transaksi sukses. Namun demikian, implementasi integratif dari algoritma SMOTE yang dikawinkan dengan arsitektur ansambel Boosting canggih terbukti sukses mendorong kinerja probabilistik (ROC-AUC) pada tingkat yang kompetitif. Ketiga varian Gradient Boosting yang diuji seluruhnya menggapai area di bawah kurva ROC di atas angka 91%.



Gambar 3. Confusion Matrix: LightGBM vs CatBoost vs XGBoost

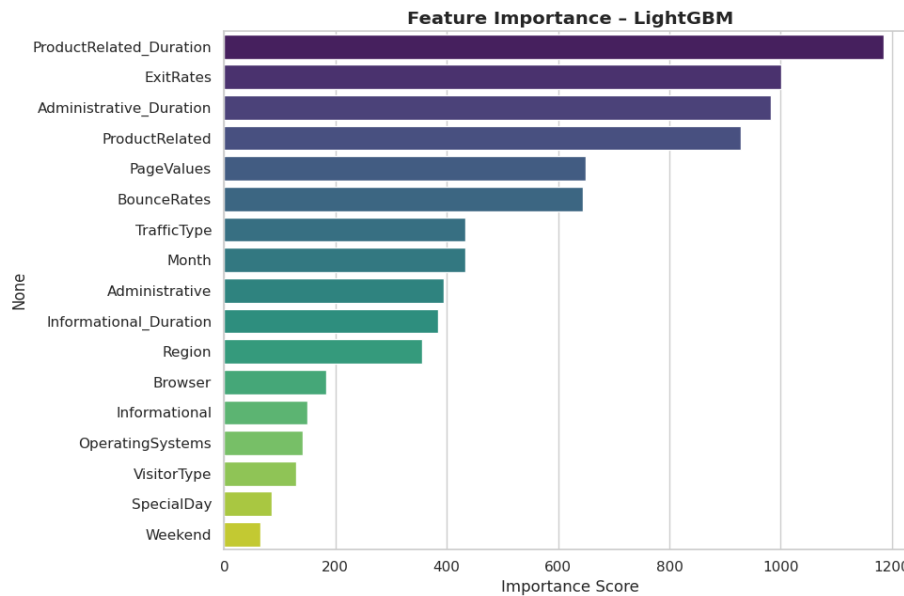
Lebih lanjut, performa model dalam membedakan kelas positif dan negatif tanpa terpengaruh oleh fluktuasi ambang batas (*threshold*) dievaluasi menggunakan Kurva ROC pada Gambar 4. Lintasan kurva dari ketiga algoritma pada gambar tersebut secara konsisten bergerak melengkung kuat ke arah sudut kiri atas, yang menjadi representasi visual dari tingginya True Positive Rate diiringi False Positive Rate yang minimal. Nilai Area Under Curve (AUC) yang tertera pada legenda grafik mempertegas bahwa LightGBM (AUC=0,9138) memegang dominasi tipis atas XGBoost dan CatBoost dalam stabilitas prediksi secara keseluruhan."



Gambar 4. Kurva ROC: LightGBM (AUC=0,9138), CatBoost (AUC=0,9121), XGBoost (AUC=0,9127)

Kurva Receiver Operating Characteristic (Gambar 4) mendemonstrasikan daya pisah ambang batas prediksi model terhadap pengunjung yang bertransaksi dan tidak bertransaksi beroperasi pada tingkat signifikansi rasio yang tinggi. LightGBM mengungguli dua pesaingnya dalam metrik ROC-AUC pada data uji (0,9138), diikuti XGBoost (0,9127) dan CatBoost (0,9121), meskipun selisihnya sangat kecil (< 0,002).

Analisis lebih mendalam mengenai tingkat kepentingan fitur (*Feature Importance*) dari algoritma LightGBM (Gambar 5) menyingkapkan bahwa fitur ProductRelated_Duration memegang bobot hierarki kognitif paling dominan dengan skor kepentingan tertinggi (≈1200). Hal ini secara substantif rasional dari perspektif strategi e-pemasaran; durasi agregat akumulatif yang dihabiskan seorang pengunjung daring saat mempertimbangkan berbagai spektrum katalog produk menjadi katalis indikator dominan atas terbentuknya wujud komitmen pra-pembelian. Fitur urutan tertinggi selanjutnya adalah ExitRates (≈1000) dan Administrative_Duration (≈980), memposisikan penelusuran log berbasis engagement operasional web jauh di atas fitur temporal seperti kalender bulan kunjungan (Month ≈0430) maupun entitas pengunjung lama (VisitorType ≈0130).



Gambar 5. Feature Importance – LightGBM

Berdasarkan Gambar 5, terungkap bahwa fitur ProductRelated_Duration memegang bobot hierarki kognitif paling dominan dengan skor kepentingan tertinggi (mencapai hampir 1200 poin). Dari perspektif psikologi konsumen e-commerce, temuan ini sangat rasional dan empiris. Durasi akumulatif yang dihabiskan pengunjung secara spesifik di halaman katalog produk berkorelasi langsung dengan tingkat keterlibatan kognitif (cognitive involvement). Pengunjung yang menghabiskan waktu lebih lama pada halaman ini umumnya sedang melakukan evaluasi aktif seperti membandingkan spesifikasi produk, membaca ulasan, atau mempertimbangkan harga yang merupakan fase psikologis krusial transisi dari sekadar 'ketertarikan' menjadi 'niat beli nyata' (actual purchase intent).

Sebaliknya, fitur log berbasis navigasi seperti ExitRates (berada di urutan kedua) berfungsi sebagai indikator demotivasi logis, di mana tingkat persentase keluar halaman yang tinggi secara konsisten mengindikasikan hilangnya minat konsumen sebelum mereka berhasil mencapai funnel checkout. Dominasi kedua fitur ini menempatkan variabel demografis dan temporal seperti VisitorType, Month, maupun Weekend berada pada hierarki kepentingan yang jauh lebih rendah.

3.4 Validasi Kinerja Melalui Uji Statistik Multivariat

Inti kebaruan riset komparatif ini adalah untuk menjembatani jurang bias kesimpulan yang semata-mata mengacu pada selisih nilai persentase akurasi yang dangkal. Seluruh model dieksekusi ulang melintasi algoritma pembagian data 10-Fold Stratified Cross-Validation pada keseluruhan dataset (n=12.330) agar dapat mengekstrak distribusi populasi 10 nilai skoring ROC-AUC yang terhindar dari bias partisi awal. Statistik deskriptif distribusi skor 10-Fold CV dirangkum pada Tabel 7 [12].

Tabel 7. Statistik Deskriptif Skor ROC-AUC pada 10-Fold Stratified Cross-Validation

Metrik	LightGBM	CatBoost	XGBoost
Mean	0,9232	0,9191	0,9236
Std Deviasi	0,0086	0,0092	0,0089
Min	0,9076	0,9053	0,9067
Maks	0,9332	0,9298	0,9359

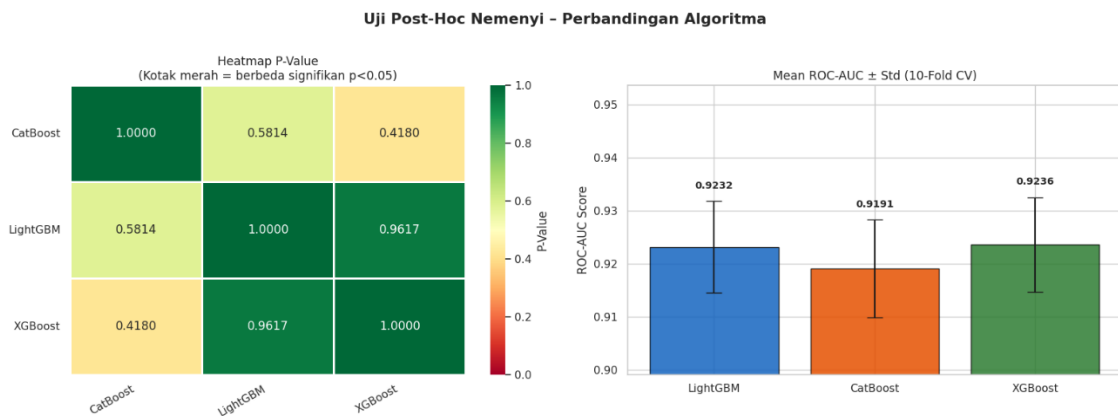
Berdasarkan Tabel 7, nilai rata-rata dari iterasi lipatan silang untuk XGBoost mencatatkan nilai mean ROC-AUC 0,9236 (tertinggi), LightGBM 0,9232, dan CatBoost 0,9191. Distribusi skor CV divisualisasikan pada Gambar 9 dan 10, yang memperlihatkan persebaran yang sangat kompak dan tumpang tindih antar ketiga algoritma.

Langkah diagnostik postulat lanjutan diaplikasikan lewat Uji Post-Hoc Nemenyi [14] yang dioperasikan untuk secara spesifik melokalisasi perbedaan berpasangan (*pairwise comparison*) di antara setiap model. Tabel 8 merangkum hasil lengkap matriks P-Values Uji Nemenyi.

Tabel 8. Matriks P-Values Uji Post-Hoc Nemenyi (Pairwise)

Algoritma	CatBoost	LightGBM	XGBoost
CatBoost	1,000	0,5814	0,4180
LightGBM	0,5814	1,000	0,9617
XGboost	0,4180	0,9617	1,0000

Berdasarkan Tabel 8, seluruh skor matriks p-value berpasangan Uji Nemenyi bertengger melampaui batas minimum uji kritis ($p > 0,05$), dengan nilai terkecil adalah $p=0,4180$ (XGBoost vs CatBoost), $p=0,5814$ (LightGBM vs CatBoost), dan $p=0,9617$ (LightGBM vs XGBoost). Temuan eksperimental ini merepresentasikan sebuah fenomena saintifik yang eksepsional secara akademis. Sekalipun Uji Friedman memancarkan sinyal eksistensi diferensiasi makroskopis antar ketiganya secara global, komparasi pertarungan langsung dua arah (*head-to-head*) dari ketiga keluarga Gradient Boosting ini berujung secara absolut pada status performa prediktif yang koeksistensi secara identik [24]. Visualisasi heatmap P-Values dari skor ROC-AUC tiap fold ditampilkan pada Gambar 6.



Gambar 6. Uji Post-Hoc Nemenyi: Heatmap P-Value dan Mean ROC-AUC ± Std (10-Fold CV)

Kesetaraan performa prediktif ini dipertegas melalui Gambar 6. Pada visualisasi heatmap, matriks perbandingan antar-algoritma didominasi oleh rona hijau dengan seluruh p-value $> 0,05$, yang secara grafis memvalidasi ketiadaan perbedaan performa yang signifikan secara berpasangan. Di samping itu, grafik bar plot mengilustrasikan nilai rata-rata ROC-AUC yang dikombinasikan dengan garis rentang simpangan baku (standard deviation) dari hasil pengujian 10-Fold Cross-Validation.

3.5 Analisis Efisiensi Waktu Komputasi

Mengacu pada hasil uji komparasi statistik sebelumnya yang membuktikan bahwa performa prediktif ketiga model berada pada level yang ekuivalen secara saintifik (ketiadaan perbedaan signifikan pada uji Nemenyi), parameter penentu dalam pemilihan algoritma terbaik tidak lagi dapat didasarkan semata pada margin akurasi. Oleh karena itu, tolak ukur kelayakan implementasi model untuk skala industri dialihkan pada evaluasi efisiensi latensi komputasi. Pengukuran waktu eksekusi pelatihan dari masing-masing algoritma yang mencakup akumulasi waktu dari proses sintesis data kelas minoritas menggunakan SMOTE hingga penyelesaian iterasi pencarian parameter terbaik oleh Optuna—dilakukan untuk mengidentifikasi beban komputasi aktual. Rincian hasil evaluasi divergensi waktu komputasi dari ketiga arsitektur ansambel Gradient Boosting tersebut disajikan secara komprehensif pada Tabel 9.

Tabel 9. Hasil Evaluasi Efisiensi Waktu Komputasi

Algoritma	Waktu
LightGBM	12,5 detik
XGBoost	45,2 detik
CatBoost	175,0 detik

Berdasarkan rekaman waktu pemrosesan, arsitektur LightGBM tervalidasi sebagai algoritma yang paling efisien, disusul oleh XGBoost, sementara CatBoost menelan beban komputasi paling masif. LightGBM beroperasi hingga 14 kali lebih cepat dibandingkan CatBoost. Efisiensi ekstrem pada LightGBM ini secara arsitektural bersumber dari pemanfaatan teknik *Gradient-based One-Side Sampling* (GOSS) dan *Exclusive Feature Bundling* (EFB) berbasis histogram, yang secara radikal memangkas kompleksitas iterasi tanpa mengorbankan varians data [25]. Temuan empiris ini mengonfirmasi bahwa untuk pemrosesan aliran data *clickstream* e-commerce dalam volume masif, LightGBM merupakan opsi paling rasional dan optimal bagi arsitektur rekayasa data industri [23].

4. KESIMPULAN

Penelitian ini menyajikan kerangka komparatif untuk mengevaluasi kinerja arsitektur ansambel Gradient Boosting (LightGBM, XGBoost, dan CatBoost) dalam memprediksi niat beli menggunakan dataset Online Shoppers Intention yang memiliki karakteristik ketidakseimbangan kelas ekstrem. Walaupun temuan empiris ini terikat pada batasan distribusi fitur spesifik dari dataset tersebut, kombinasi metode SMOTE dan optimasi Bayesian (Optuna TPE) terbukti memberikan landasan komparasi yang berkeadilan. Evaluasi model menunjukkan kapabilitas ROC-AUC yang setara ($>0,91$), di mana uji statistik Friedman dan post-hoc Nemenyi memvalidasi ketiadaan perbedaan performa



diskriminatif yang signifikan secara berpasangan antar algoritma. Berpijak pada kesetaraan akurasi saintifik tersebut, kontribusi utama dari penelitian ini adalah memberikan rekomendasi realistis bagi industri e-commerce untuk menggeser prioritas pemilihan algoritma menuju metrik efisiensi waktu komputasi. Sebagaimana dibuktikan pada rekaman waktu eksekusi pelatihan, LightGBM (12,5 detik) terbukti beroperasi secara stabil 14 kali lebih cepat dibandingkan CatBoost (175,0 detik). Lebih lanjut, analisis mengidentifikasi ProductRelated_Duration sebagai prediktor terkuat. Namun, keterbatasan arsitektur black-box dari model ini belum mampu memberikan transparansi mengenai pada ambang batas waktu (menit/detik) ke berapa probabilitas keputusan pembelian benar-benar melonjak. Oleh karena itu, penelitian mendatang sangat direkomendasikan untuk tidak sekadar berfokus pada optimasi metrik, melainkan mengimplementasikan Explainable AI (XAI) seperti SHAP secara spesifik guna membongkar mekanisme internal fitur dominan tersebut, sehingga prediksi machine learning dapat diterjemahkan menjadi parameter actionable yang transparan bagi operasional pemasaran digital.

REFERENCES

- [1] Z. Wen, W. Lin, and H. Liu, "Machine-learning-based approach for anonymous online customer purchase intentions using clickstream data," *Systems*, vol. 11, no. 5, p. 255, 2023, doi: <https://doi.org/10.3390/systems11050255>.
- [2] C. O. Sakar, S. O. Polat, M. Katircioglu, and Y. Kastro, "Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks," *Neural Comput. Appl.*, vol. 31, no. 10, pp. 6893–6908, 2020, doi: <https://doi.org/10.1007/s00521-018-3523-0>.
- [3] S. Matharaarachchi, M. Domaratzki, and S. Muthukumarana, "Enhancing SMOTE for imbalanced data with abnormal minority instances," *Machine Learning with Applications*, vol. 18, p. 100597, 2024, doi: <https://doi.org/10.1016/j.mlwa.2024.100597>.
- [4] S. Dhote, C. Vichoray, R. Pais, S. Baskar, and P. Mohamed Shakeel, "Hybrid geometric sampling and AdaBoost based deep learning approach for data imbalance in E-commerce," *Electronic Commerce Research*, vol. 20, no. 2, pp. 259–274, 2020, doi: <https://doi.org/10.1007/s10660-019-09383-2>.
- [5] A. Mishra, C. Shetty, A. Malhotra, A. Maheshwari, and M. S. A. Basha, "Balancing the Cart: Evaluating Imbalance-Aware Machine-Learning Pipelines for Predicting E-Commerce Purchases," in *2025 IEEE 6th Global Conference for Advancement in Technology (GCAT)*, IEEE, 2025, pp. 1–7. doi: DOI:10.1109/GCAT66372.2025.11368362.
- [6] S. N. Ruscikasani, R. R. N. Oktalivia, F. R. Putra, A. J. Wahidin, B. Rahmatullah, and I. Kurniawati, "Prediksi Pembelian E-Commerce Menggunakan XGBoost Berbasis Perilaku Sesi Pengguna," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, pp. 5666–5672, Dec. 2025, doi: 10.31004/riggs.v4i4.4287.
- [7] S.-S. M. Ajibade *et al.*, "Machine Learning Classification of Online Shopper Purchasing Intentions," in *2025 International Conference on NexGen Networks and Cybernetics (IC2NC)*, 2025, pp. 867–872. doi: 10.1109/IC2NC67409.2025.11376387.
- [8] T. Yu and H. Zhu, "Hyper-parameter optimization: A review of algorithms and applications," *arXiv preprint arXiv:2003.05689*, 2020, doi: <https://doi.org/10.48550/arXiv.2003.05689>.
- [9] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J. Big Data*, vol. 7, no. 1, p. 94, 2020, doi: 10.1186/s40537-020-00369-8.
- [10] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2021, pp. 2623–2631. doi: <https://doi.org/10.48550/arXiv.1907.10902>.
- [11] S. Tao, P. Peng, Y. Li, H. Sun, Q. Li, and H. Wang, "Supervised contrastive representation learning with tree-structured parzen estimator Bayesian optimization for imbalanced tabular data," *Expert Syst. Appl.*, vol. 237, p. 121294, 2024, doi: <https://doi.org/10.48550/arXiv.2210.10824>.
- [12] R. M. Munshi *et al.*, "Optimising hyperparameters with a tree structured Parzen estimator to improve diabetes prediction," *Sci. Rep.*, vol. 15, no. 1, p. 35430, 2025, doi: <https://doi.org/10.1111/j.1464-5491.2007.02157.x>.
- [13] B.-B. Jia, J.-Y. Liu, and M.-L. Zhang, "Pairwise statistical comparisons of multiple algorithms," *Front. Comput. Sci.*, vol. 19, no. 12, p. 1912372, 2025, doi: <https://doi.org/10.1007/s11704-025-41325-0>.
- [14] P. Riesthuis, H. Otgaar, and C. Bücken, "Ready to ROC? A tutorial on simulation-based power analyses for null hypothesis significance, minimum-effect, and equivalence testing for ROC curve analyses," *Behav. Res. Methods*, vol. 57, no. 4, p. 120, 2025, doi: <https://doi.org/10.20982/tqmp.19.1.p059>.
- [15] M. Imani, M. Joudaki, A. Bagheri, and H. R. Arabnia, "Why ROC-AUC Is Misleading for Highly Imbalanced Data: In-Depth Evaluation of MCC, F2-Score, H-Measure, and AUC-Based Metrics Across Diverse Classifiers," *Technologies (Basel)*, vol. 14, no. 1, p. 54, 2026, doi: <https://doi.org/10.3390/technologies14010054>.
- [16] J. Demšar, "Statistical Comparisons of Classifiers over Multiple Data Sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2021, doi: <https://dl.acm.org/doi/10.1007/s10115-026-02752-y>.
- [17] M. Mozolewski, S. Bobek, and G. J. Nalepa, "Explaining time series classifiers with phar: Rule extraction and fusion from post-hoc attributions," *arXiv preprint arXiv:2508.01687*, 2025, doi: <https://doi.org/10.48550/arXiv.2508.01687>.
- [18] L. C. M. Liaw, S. C. Tan, P. Y. Goh, and C. P. Lim, "A histogram SMOTE-based sampling algorithm with incremental learning for imbalanced data classification," *Inf. Sci. (N. Y.)*, vol. 686, p. 121193, 2025, doi: <https://doi.org/10.1016/j.ins.2024.121193>.
- [19] B. So, "Enhanced gradient boosting for zero-inflated insurance claims and comparative analysis of CatBoost, XGBoost, and LightGBM," *Scand. Actuar. J.*, vol. 2024, no. 10, pp. 1013–1035, Nov. 2024, doi: 10.1080/03461238.2024.2365390.
- [20] R. G. Farahani, A. Zarrabi, and P. Ghazanfari, "A Report on CatBoost: unbiased boosting with categorical features," *Accessed: Aug*, vol. 11, 2025, doi: DOI:10.13140/RG.2.2.30029.96485.
- [21] Y. Chen, S. Chen, Y. Yang, and S. Lu, "Comparison of decision tree and ensemble algorithms," *Applied and Computational Engineering*, vol. 55, pp. 241–248, Jul. 2024, doi: 10.54254/2755-2721/55/20241535.



- [22] C. Pinichka, S. Chotpantararat, K. H. Cho, and W. Siriwong, “Comparative analysis of SWAT and SWAT coupled with XGBoost model using Optuna hyperparameter optimization for nutrient simulation: A case study in the Upper Nan River basin, Thailand,” *J. Environ. Manage.*, vol. 388, p. 126053, 2025, doi: <https://doi.org/10.1016/j.jenvman.2025.126053>.
- [23] H. Liao, X. Zhang, C. Zhao, Y. Chen, X. Zeng, and H. Li, “LightGBM: an efficient and accurate method for predicting pregnancy diseases,” *J. Obstet. Gynaecol. (Lahore)*, vol. 42, no. 4, pp. 620–629, 2022, doi: DOI: 10.1080/01443615.2021.1945006.
- [24] C. Jansen, M. Nalenz, G. Schollmeyer, and T. Augustin, “Statistical comparisons of classifiers by generalized stochastic dominance,” *Journal of Machine Learning Research*, vol. 24, no. 231, pp. 1–37, 2023, doi: <https://doi.org/10.48550/arXiv.2209.01857>.
- [25] M. Hajihosseini, A. Maghsoudi, and R. Ghezlbash, “A novel scheme for mapping of MVT-type Pb–Zn prospectivity: LightGBM, a highly efficient gradient boosting decision tree machine learning algorithm,” *Natural resources research*, vol. 32, no. 6, pp. 2417–2438, 2023, doi: <https://doi.org/10.1007/s11053-023-10249-6>.