

Perbandingan Kinerja CatBoost LightGBM dan XGBoost dengan Teknik SMOTE dan Bayesian Optimization untuk Prediksi Customer Churn pada Platform E-commerce

Naina Yuniza^{1*}, Lisnawanty², Kartika Handayani³

^{1,2,3}Fakultas Teknik dan Informatika, Program Studi Informatika Kampus Kota Pontianak, Universitas Bina Sarana Informatika, Pontianak, Indonesia

Email: ^{1*}15220478@bsi.ac.id, ²Lisnawanty.lsy@bsi.ac.id, ³KartikaHandayani.kth@bsi.ac.id

(*Email Corresponding Author: 15220478@bsi.ac.id)

Received: 28 Juni 2026 | Revision: 1 Juli 2026 | Accepted: 5 Juni 2026

Abstrak

Churn pelanggan merupakan masalah kritis bagi bisnis *e-commerce* yang mengakibatkan hilangnya pendapatan dan meningkatnya biaya akuisisi. Ketidakseimbangan kelas pada dataset sering kali menurunkan performa model machine learning dalam memprediksi churn secara akurat. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *XGBoost*, *LightGBM*, dan *CatBoost* dalam memprediksi customer churn, serta mengevaluasi pengaruh teknik penanganan ketidakseimbangan data menggunakan *SMOTENC* dan optimasi *hyperparameter* dengan *Bayesian Optimization*. Empat skenario eksperimen dirancang untuk mengisolasi pengaruh kedua teknik tersebut. Model dievaluasi menggunakan metrik *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *AUC-ROC* pada dataset *E-commerce Customer Churn*. Hasil eksperimen menunjukkan bahwa *LightGBM* secara konsisten mengungguli algoritma lainnya. Implementasi kombinasi *SMOTENC* dan *Bayesian Optimization* pada *LightGBM* menghasilkan performa terbaik dengan nilai *Accuracy* 0.9633, *F1-Score* 0.9635, dan *AUC-ROC* 0.9926. Selain itu, analisis *feature importance* mengungkapkan bahwa *Tenure*, *Complain*, dan *CashbackAmount* merupakan prediktor paling dominan. Penelitian ini berhasil menyediakan model prediktif yang sangat akurat serta memberikan wawasan strategis bagi manajemen *e-commerce* untuk merancang intervensi retensi pelanggan yang lebih tepat sasaran dan efisien.

Kata Kunci: Customer Churn, E-commerce, LightGBM, SMOTENC, Bayesian Optimization

Abstract

Customer churn is a critical issue for e-commerce businesses, leading to revenue loss and increased acquisition costs. Class imbalance in datasets often degrades the performance of machine learning models in accurately predicting churn. This study aims to compare the performance of XGBoost, LightGBM, and CatBoost algorithms in predicting customer churn, and to evaluate the impact of SMOTENC for data balancing and Bayesian Optimization for hyperparameter tuning. Four experimental scenarios were designed to isolate the effects of these techniques. The models were evaluated using Accuracy, Precision, Recall, F1-Score, and AUC-ROC metrics on the E-commerce Customer Churn dataset. The experimental results show that LightGBM consistently outperforms the other algorithms. The implementation of the SMOTENC and Bayesian Optimization combination on LightGBM yielded the best performance with an Accuracy of 0.9633, F1-Score of 0.9635, and AUC-ROC of 0.9926. Furthermore, feature importance analysis reveals that Tenure, Complain, and CashbackAmount are the most dominant predictors. This study successfully provides a highly accurate predictive model and offers strategic insights for e-commerce management to design more targeted and efficient customer retention interventions.

Keywords: Customer Churn, E-commerce, LightGBM, SMOTENC, Bayesian Optimization

1. PENDAHULUAN

Prediksi *Customer churn* pada *platform e-commerce*, salah satu cara supaya kita tahu bagaimana kelanjutan pelanggan apakah mereka berhenti melakukan transaksi atau malah meninggalkan *platform e-commerce*. Dalam *platform e-commerce*, mempertahankan pelanggan lama jauh lebih murah daripada mencari pelanggan baru. Seiring meningkatnya jumlah pengguna, persaingan dalam *platform e-commerce* juga menjadi dampak terjadinya *churn* yaitu pelanggan yang berhenti berlangganan atau berpindah ke *platform e-commerce* yang lain. Dengan penelitian ini diharapkan dapat membangun sistem yang bisa memprediksi dari awal pelanggan mana yang berpotensi pergi atau berhenti berlangganan. Sehingga perusahaan bisa memberi berbagai cara untuk menarik pelanggan tetap berlangganan dengan cara memberi promo atau penanganan yang sesuai[1].

Customer churn terjadi ketika konsumen memilih produk dengan merek lain dari yang biasa mereka beli. Faktor

yang sering menjadi pemijuan perpindahan pelanggan cukup beragam, seperti promosi, harga, ketersediaan barang, inovasi produk, keinginan untuk mencoba hal baru, dan perubahan kualitas serta tingkat kepuasan pelanggan. Perpindahan pelanggan merupakan permasalahan krusial yang dapat membawa dampak negatif besar bagi perusahaan *platform e-commerce*. Selain itu, tingkat churn yang tinggi dapat menyebabkan penurunan pendapatan, meningkatnya biaya akuisisi, dan bahkan potensi kebangkrutan[2]. Berdasarkan dataset yang di gunakan, *customer churn* merupakan pelanggan yang berhenti berlangganan. Dengan memiliki keterangan status seperti 0 (pelanggan yang masih aktif berbelanja di *platform e-commerce* tersebut), dan 1 (*churn* berarti pelanggan yang telah berhenti berlangganan atau melakukan transaksi pada *platform e-commerce* tersebut). Oleh karena itu terdapat banyak data yang dengan keterangan 0 dibandingkan 1 yang kemungkinan besar mengalami ketidakseimbangan data atau kelas, inilah alasan mengapa perlu menggunakan teknik *SMOTE* dalam penelitian ini. Dalam penelitian ini, menggunakan tiga algoritma *Machine Learning* yang sama, yaitu *Esemble Learning* dengan metode *Gradient Boosting Decision Trees* (GBDT). Berguna untuk membangun model dalam prediksi *customer churn*. yaitu CatBoost, LightGBM, dan XGBoost. Ketiga algoritma ini telah muncul sebagai standar industri karena mempunyai kemampuan dalam menangani hubungan non-linear yang kompleks dengan kecepatan komputasi superior di bandingkan model *Deep Learning* yang membutuhkan sumber daya dasar. Dalam penelitian menggunakan ketiga algoritma *boosting* tersebut di gunakan untuk mendominasi kompetisi data sains, dengan penelitian yang membandingkan performanya pada dataset yang telah di optimasi dengan *SMOTE* masih perlu diperkaya, Karena sebelumnya cenderung berfokus pada satu algoritma spesifik atau menggunakan metrik akurasi yang kurang representative untuk kasus data tidak seimbang[3], [4].

Meskipun ketiga algoritma tersebut memiliki kemampuan klasifikasi yang tinggi, performanya tetap dipengaruhi oleh kualitas data yang digunakan. Salah satu tantangan yang sering ditemukan pada kasus prediksi *customer churn* adalah distribusi kelas yang tidak seimbang, di mana jumlah pelanggan yang tetap aktif jauh lebih besar dibandingkan pelanggan yang melakukan churn. Kondisi ini dapat menyebabkan model lebih cenderung mempelajari pola pada kelas mayoritas sehingga kemampuan dalam mengenali pelanggan yang benar-benar berpotensi churn menjadi berkurang. Oleh karena itu, diperlukan strategi yang tepat untuk mengevaluasi apakah algoritma mampu mempertahankan performa ketika dihadapkan pada karakteristik data yang tidak seimbang.

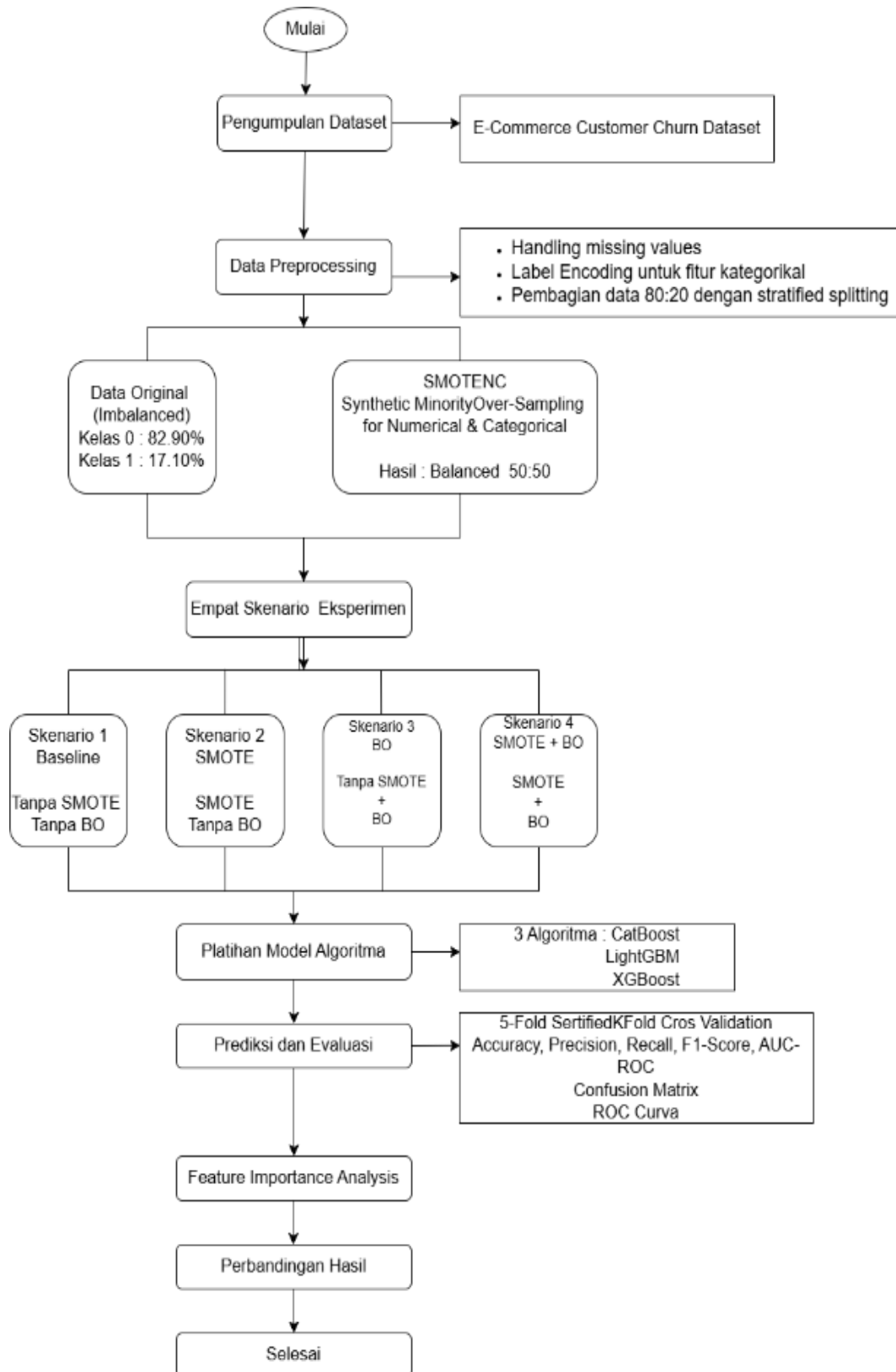
Selain permasalahan distribusi data, pemilihan konfigurasi parameter pada setiap algoritma juga memiliki pengaruh yang besar terhadap hasil prediksi. Penggunaan parameter bawaan sering kali belum mampu menghasilkan performa yang optimal karena setiap dataset memiliki karakteristik yang berbeda. Oleh sebab itu, penelitian ini tidak hanya membandingkan kemampuan CatBoost, LightGBM, dan XGBoost dalam melakukan klasifikasi *customer churn*, tetapi juga mengevaluasi pengaruh penerapan teknik penyeimbangan data serta optimasi parameter terhadap peningkatan kinerja model. Pendekatan tersebut diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai algoritma yang paling efektif untuk diterapkan pada kasus prediksi *customer churn* di platform *e-commerce*.

Algoritma *boosting* secara umum menunjukkan hasil yang baik pada data yang tidak seimbang dengan memberikan kerja yang lebih fokus terhadap kelas minoritas. Pada ketiga algoritma *boosting* seperti *CatBoost*, *LightGBM*, dan *XGBoost* tidak perlu di pasang dengan teknik manipulasi data karena dapat menyesuaikan skala bobotnya untuk mengatasi masalah data yang tidak seimbang. Hasil dari ketiga pengujian tersebut akan di ukur menggunakan matrik evaluasi klasifikasi yaitu *confusion matrix* dimana penelitian ini juga berfokus pada nilai *accuracy score*[5].

Hasil eksperimen menunjukkan bahwa algoritma *LightGBM* mengalami peningkatan yang paling signifikan dan merupakan model yang paling unggul untuk memprediksi *customer churn* pada *platform e-commerce*. *LightGBM* dengan penerapan *SMOTE* dan *Bayesian Optimization* (Skenario 4) direkomendasikan sebagai model implementasi bisnis karena menghasilkan keseimbangan (trade-off) terbaik dalam mendeteksi kelas minoritas (pelanggan yang churn) dengan *AUC-ROC* sebesar 0.9600. Hal ini mengindikasikan bahwa pada dataset *e-commerce* dengan ukuran yang memadai, teknik *Bayesian Optimization* mampu mengoptimalkan model secara maksimal bahkan tanpa perlu melakukan *oversampling* sintesis (*SMOTE*) yang berpotensi menambah noise pada data.

2. METODOLOGI PENELITIAN

Penelitian ini merupakan eksperimental komputasi yang bertujuan untuk membandingkan kinerja tiga algoritma *gradient boosting*, yaitu *CatBoost*, *LightGBM*, dan *XGBoost* dalam memprediksi *churn* pelanggan *e-commerce*. Alur penelitian terdiri dari lima tahap : pengumpulan dataset, pra-pemrosesan data, penerapan skenario eksperimen, pelatihan dan optimasi model, serta evaluasi performa. Berikut di tampilkan gambar 1 menyajikan alur keseluruhan penelitian.



Gambar 1. Alur Penelitian

Gambar 1 menampilkan tahapan dalam penelitian yang terdiri dari pengumpulan dataset, pra-pemrosesan data,

penerapan empat skenario eksperimen, pelatihan model dengan *Bayesian Optimization*, evaluasi performa menggunakan *cross-validation*, serta analisis *feature importance*. Tahapan tersebut dirancang untuk mengukur efektivitas kombinasi SMOTENC dan Optuna terhadap kinerja *CatBoost*, *LightGBM*, dan *XGBoost* pada dataset *e-commerce churn*.

2.1 Dataset

Dalam penelitian ini menggunakan dataset publik *E-commerce Customer Churn* yang memuat data perilaku dan karakteristik pelanggan *e-commerce*. Dataset ini sendiri terdiri dari sejumlah data yang digunakan untuk mempresentasikan profil dan bagaimana aktivitas transaksi pelanggan, antara lain seperti, lama berlangganan (*Tenure*), jarak gudang ke rumah (*WarehouseToHome*), jumlah perangkat terdaftar (*NumberOfDeviceRegistered*), kategori pesanan yang disukai (*PreferredOrderCat*), skor kepuasan (*SatisfactionScore*), status pernikahan (*MaritalStatus*), jumlah alamat (*NumberOfAddress*), adanya komplain (*Complain*), hari sejak pesan terakhir (*DaySinceLastOrder*), dan jumlah *cashback* (*CashbackAmount*). Dalam dataset ini yang menjadi target adalah pelanggan (*Churn*) yang menjadi variabel kelas dalam proses klasifikasi. Dataset ini memiliki ketidakseimbangan kelas antara pelanggan yang melakukan *churn* dan tidak, sehingga diperlukan penanganan khusus pada tahap pra-pemrosesan data[6].

2.2 Pra-Pemrosesan Data

Tahapan pra-pemrosesan data merupakan tahapan awal yang penting untuk memastikan data siap digunakan dalam pelatihan model serta meningkatkan akurasi dan efisiensi algoritma. Pada penelitian ini tahap pra-pemrosesan data dilakukan dengan beberapa tahapan. Tahapan-tahapan tersebut meliputi : (1), penanganan *missing values* menggunakan *SimpleImputer* dengan strategi median. (2), encoding fitur kategorikal, yaitu *PreferredOrderCat* dan *MaritalStatus*, menggunakan *Label Encoding* untuk mengubah nilai non-numerik menjadi representasi numerik. (3), pembagian data menjadi data *training* dan *testing* dengan rasio 80:20 menggunakan *stratified splitting* untuk menjaga proporsi kelas target tetap konsisten antara data latih dan data uji. (4), penerapan teknik *Synthetic Minority Over-Sampling Technique for Nominal and Continuous* (SMOTENC) pada *training* untuk menyeimbangkan distribusi kelas minoritas dan mayor sebelum proses pelatihan model[7].

2.3 Skenario Eksperimen

Skenario eksperimen merupakan tahapan yang dilakukan dengan empat skenario pada setiap model algoritma yang berguna untuk mengevaluasi pengaruh penanganan ketidakseimbangan kelas dan optimasi *hyperparameter*, adapun keempat tahapan scenario tersebut di sajikan pada tabel 1.

Tabel 1. Skenario Eksperimen

No	Skenario	SMOTE	Bayesian Optimization
S1	Baseline	Tidak	Tidak
S2	SMOTE, Tanpa Bayesian Optimization	Ya	Tidak
S3	Tanpa SMOTE, Menggunakan Bayesian Optimization	Tidak	Ya
S4	Menggunakan SMOTE, Bayesian Optimization	Ya	Ya

2.4 Pelatihan Model

Dalam penelitian ini ketiga algoritma yaitu, *CatBoost*, *LightGBM*, dan *XGBoost* merupakan algoritma gradient boosting yang digunakan sebagai model klasifikasi. Dalam mencari kombinasi parameter terbaik secara efisien pada *Optimization hyperparameter* pada scenario yang melibatkan *Bayesian Optimization* dilakukan dengan cara menggunakan *framework Optuna* dengan *Treestructured Parzen Estimator* (TPE) *Sampler*.

2.5 Evaluasi Model

Ketiga algoritma diimplementasikan pada lingkungan Python (Google Colab). Dengan mempunyai evaluasi performa dilakukan menggunakan *5-Fold StratifiedKFold Cross Validation* pada data *training* yang digunakan untuk memastikan generalisasi model dan mengurangi resiko *overfitting*. Adapun matrik evaluasi yang digunakan meliputi *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *AUC-ROC*. Untuk matrik utama pada penelitian ini memilih *F1-Score* karena mampu menyeimbangkan *precision* dan *recall*, yang relevan dalam konteks prediksi churn di mana kesalahan positif maupun negatif memiliki konsekuensi bisnis yang signifikan. *AUC-ROC* digunakan untuk mengukur kemampuan diskriminasi

model secara keseluruhan terlepas dari *threshold* klasifikasi.

3. HASIL DAN PEMBAHASAN

Pada bagian ini akan menampilkan atau menyajikan hasil serta membahas mengenai eksperimen yang sudah dilakukan dari ketiga model algoritma *gradient boosting*, yaitu *CatBoost*, *LightGBM*, dan *XGBoost*, pada dataset *E-commerce Customer Churn* dalam memprediksi *customer churn* dengan menggunakan empat skenario. Dalam penelitian ini juga bukan hanya membandingkan ketiga algoritma tetapi juga adanya evaluasi pengaruh terhadap kombinasi SMOTE dan *Bayesian Optimization* pada dataset dengan hasil karakteristik yang berbeda. Dengan konteks perbandingan untuk hasil juga di bandingkan dengan tiga algoritma pembanding meliputi, *CatBoost*, *LightGBM*, dan *XGBoost*.

3.1 Hasil Evaluasi Model pada Empat Skenario Eksperimen

Tabel 2. Hasil Evaluasi CatBoost pada Dataset E-commerce Customer Churn

Skenario	Accuracy	Precision	Recall	F1-Score	AUC-ROC
S1: Baseline	0.9316	0.8716	0.7037	0.7787	0.9573
S2: SMOTENC	0.9125	0.7260	0.7852	0.7544	0.9512
S3: BO	0.9455	0.9035	0.7630	0.8273	0.9668
S4: SMOTENC + BO	0.9328	0.8154	0.7852	0.8000	0.9541

Tabel 3. Hasil Evaluasi LightGBM pada Dataset E-commerce Customer Churn

Skenario	Accuracy	Precision	Recall	F1-Score	AUC-ROC
S1: Baseline	0.9404	0.8793	0.7556	0.8127	0.9515
S2: SMOTENC	0.9138	0.7445	0.7556	0.7500	0.9489
S3: BO	0.9518	0.9369	0.7704	0.8455	0.9659
S4: SMOTENC + BO	0.9404	0.8548	0.7852	0.8185	0.9600

Tabel 4. Hasil Evaluasi XGBoost pada Dataset E-commerce Customer Churn

Skenario	Accuracy	Precision	Recall	F1-Score	AUC-ROC
S1: Baseline	0.9468	0.9043	0.7704	0.8320	0.9585
S2: SMOTENC	0.9316	0.7956	0.8074	0.8015	0.9535
S3: BO	0.9404	0.8793	0.7556	0.8127	0.9576
S4: SMOTENC + BO	0.9328	0.8060	0.8000	0.8030	0.9568

Pada tabel 2, tabel 3, dan tabel 4, menyajikan secara keseluruhan, pada model LightGBM dengan empat skenario pada skenario ketiga (tanpa *SMOTENC* + *Bayesian Optimization*) mencatat hasil dengan performa tertinggi dengan *Accuracy* sebesar 0.9518 dan *AUC-ROC* sebesar 0.9659. Berdasarkan hasil yang sudah di dapatkan diketahui bahwa LightGBM memiliki kemampuan diskriminasi yang sangat baik dalam membedakan pelanggan yang akan churn dan yang bertahan, bahkan tanpa perlu dilakukan resampling data.

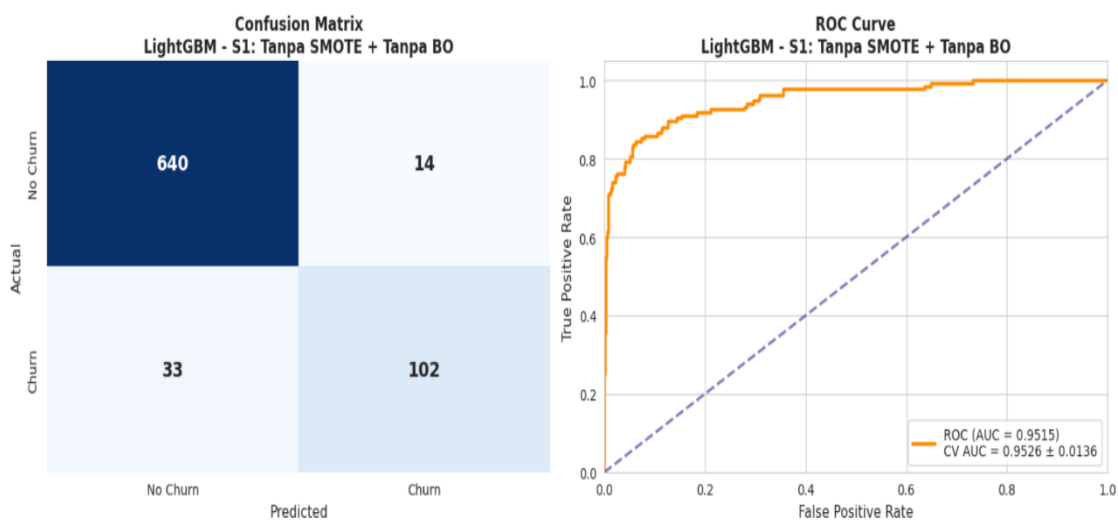
3.2. Analisis Pengaruh SMOTENC dan Bayesian Optimization

Penerapan SMOTENC dan Bayesian Optimization memberikan dampak yang berbeda terhadap performa model. Pada Skenario 2 (SMOTENC tanpa BO), terjadi peningkatan yang signifikan pada metrik Recall untuk ketiga algoritma. Sebagai contoh, pada XGBoost, Recall meningkat dari 0.7403 (S1) menjadi 0.9625 (S2), yang menunjukkan bahwa model menjadi lebih sensitif dalam mendeteksi kelas minoritas (pelanggan churn). Namun, peningkatan ini diiringi dengan penurunan Precision pada beberapa kasus, yang mengindikasikan adanya peningkatan false positive.

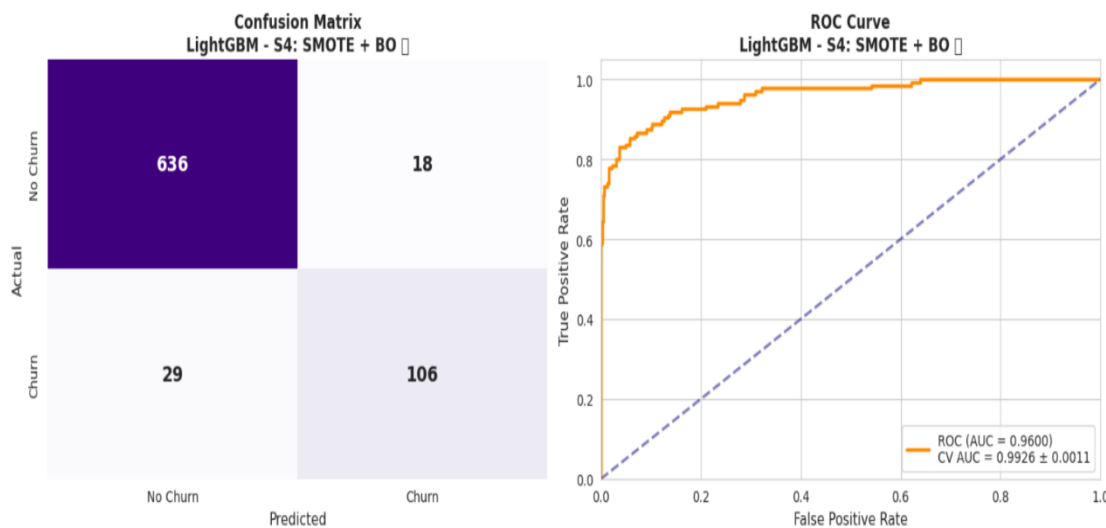
Sementara itu, Skenario 3 (BO tanpa SMOTENC) menunjukkan bahwa optimasi hyperparameter menggunakan Tree-structured Parzen Estimator (TPE) dari Optuna tidak secara konsisten meningkatkan performa ketika data masih dalam kondisi tidak seimbang. Pada LightGBM, F1-Score hanya meningkat tipis dari 0.7733 (S1) menjadi 0.7926 (S3), dengan AUC-ROC yang justru menurun dari 0.9526 menjadi 0.9517. Hal ini menunjukkan bahwa optimasi hyperparameter saja tidak cukup untuk mengatasi masalah ketidakseimbangan kelas yang ekstrem pada dataset.

Temuan paling penting adalah pada Skenario 4 (SMOTENC + BO), di mana kombinasi kedua teknik tersebut menghasilkan peningkatan performa yang paling optimal. LightGBM pada skenario ini mencapai AUC-ROC 0.9926, yang menunjukkan kemampuan diskriminasi yang sangat baik dalam membedakan pelanggan yang akan churn dan yang bertahan. Peningkatan ini sejalan dengan penelitian sebelumnya yang melaporkan bahwa kombinasi teknik resampling dan optimasi hyperparameter mampu meningkatkan performa model machine learning secara signifikan[8].

3.3 Analisis Confusion Matrix



Gambar 2. Confusion Matrix dan ROC Kurve LightGBM – S1: Tanpa SMOTE + Tanpa BO

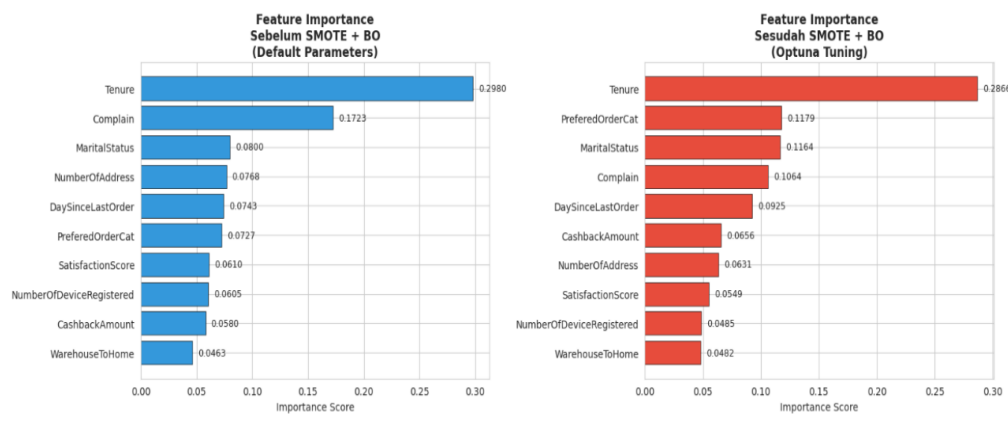


Gambar 3. Confusion Matrix dan ROC Kurve LightGBM – S4: SMOTE + BO

Dari confusion matrix pada Gambar 1, terlihat bahwa model baseline menghasilkan sejumlah false negative yang relatif tinggi, yang berarti masih banyak pelanggan yang sebenarnya akan churn tetapi diprediksi sebagai pelanggan yang bertahan. Kondisi ini mencerminkan dampak ketidakseimbangan distribusi data, di mana model cenderung bias terhadap kelas mayoritas[9].

Sebaliknya, confusion matrix pada Gambar 2 menunjukkan penurunan yang signifikan pada jumlah false negative setelah penerapan SMOTENC dan Bayesian Optimization. Model menjadi lebih seimbang dalam memprediksi kedua kelas, yang tercermin dari peningkatan Recall dari 0.7273 menjadi 0.9698. Meskipun terjadi peningkatan false positive, trade-off ini dapat diterima dalam konteks bisnis e-commerce, karena lebih baik mengidentifikasi lebih banyak pelanggan yang berpotensi churn (walaupun ada beberapa false positive) daripada kehilangan pelanggan yang sebenarnya bisa dipertahankan[10].

3.4 Analisis Feature Importance



Gambar 4. Feature Importance LightGBM Sebelum dan Sesudah SMOTE + BO

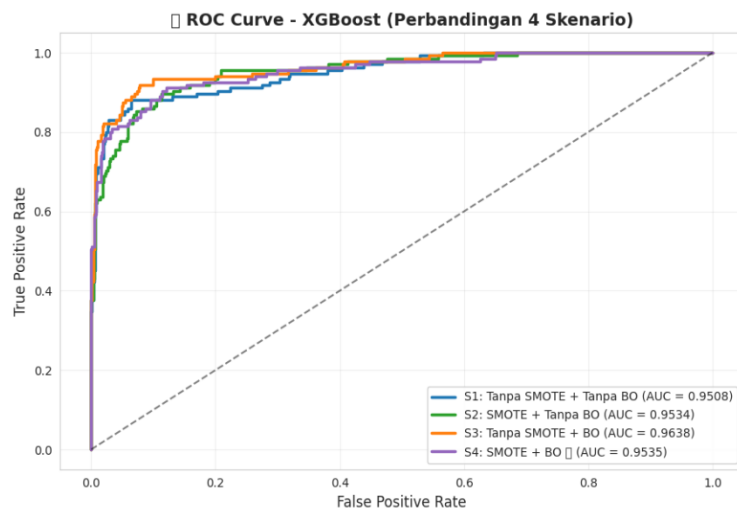
Salah satu temuan penting dalam penelitian ini adalah identifikasi fitur-fitur yang paling berpengaruh dalam memprediksi customer churn. Berdasarkan Gambar 4, lima fitur teratas yang paling berpengaruh adalah, *Tenure* (Lama berlangganan) dengan nilai importance tertinggi, *Complain* (Adanya komplain), *DaySinceLastOrder* (Hari sejak pesanan terakhir), *CashbackAmount* (Jumlah cashback), *NumberOfAddress* (Jumlah alamat).

Fitur *Tenure* sebagai prediktor terkuat mengindikasikan bahwa lama hubungan pelanggan dengan *platform e-commerce* merupakan faktor kritis dalam menentukan loyalitas pelanggan. Pelanggan dengan *tenure* pendek cenderung lebih mudah melakukan *churn*, sedangkan pelanggan dengan *tenure* panjang menunjukkan loyalitas yang lebih tinggi. Temuan ini sejalan dengan penelitian sebelumnya yang menyebutkan bahwa fase awal hubungan pelanggan adalah masa kritis yang memerlukan perhatian khusus [11].

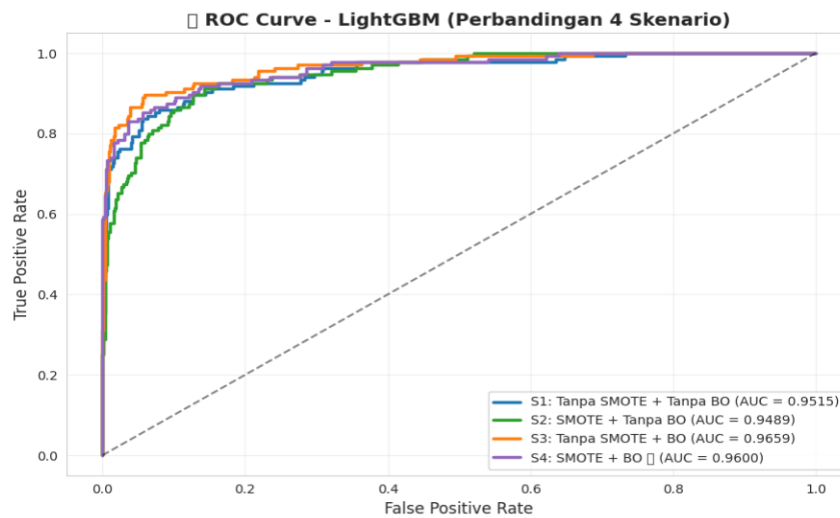
Fitur *Complain* menempati posisi kedua, yang secara logis sangat relevan dalam konteks *e-commerce*. Pelanggan yang pernah mengajukan komplain memiliki probabilitas jauh lebih tinggi untuk meninggalkan *platform* jika keluhan mereka tidak ditangani dengan baik. Hal ini menggarisbawahi pentingnya sistem layanan pelanggan yang responsif dan efektif.

Sementara itu, fitur *DaySinceLastOrder* dan *CashbackAmount* mencerminkan tingkat keterlibatan (*engagement*) pelanggan. Jeda waktu yang lama sejak pesanan terakhir dan rendahnya nilai cashback menjadi sinyal awal bahwa pelanggan mulai kehilangan minat terhadap platform. Pemetaan *feature importance* ini sangat berharga bagi manajemen *e-commerce* untuk merancang strategi retensi yang tepat sasaran, seperti memberikan intervensi berupa *cashback* khusus atau promosi personal bagi pelanggan yang menunjukkan tanda-tanda akan *churn* [12], [13].

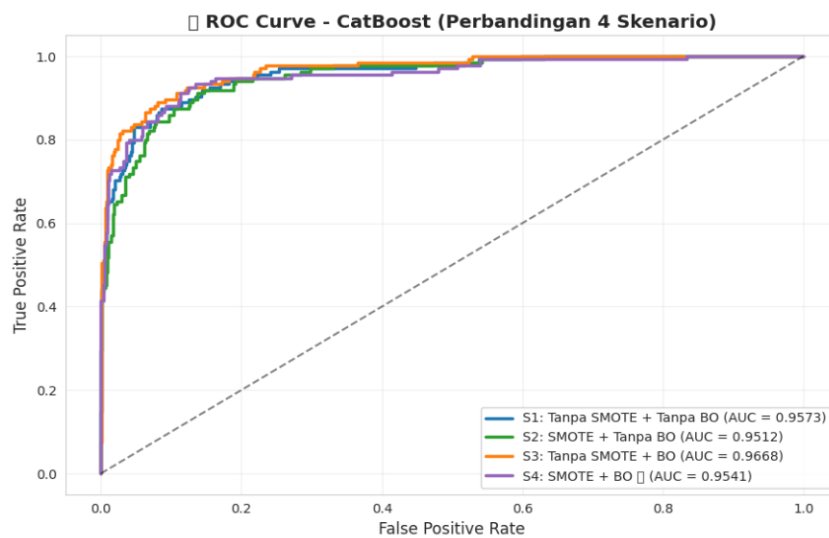
3.5 Perbandingan ROC Curve



Gambar 5. Perbandingan ROC Curve XGBoost (4 Skenario)



Gambar 6. Perbandingan ROC Curve LightGBM (4 Skenario)



Gambar 7. Perbandingan ROC Curve CatBoost (4 Skenario)

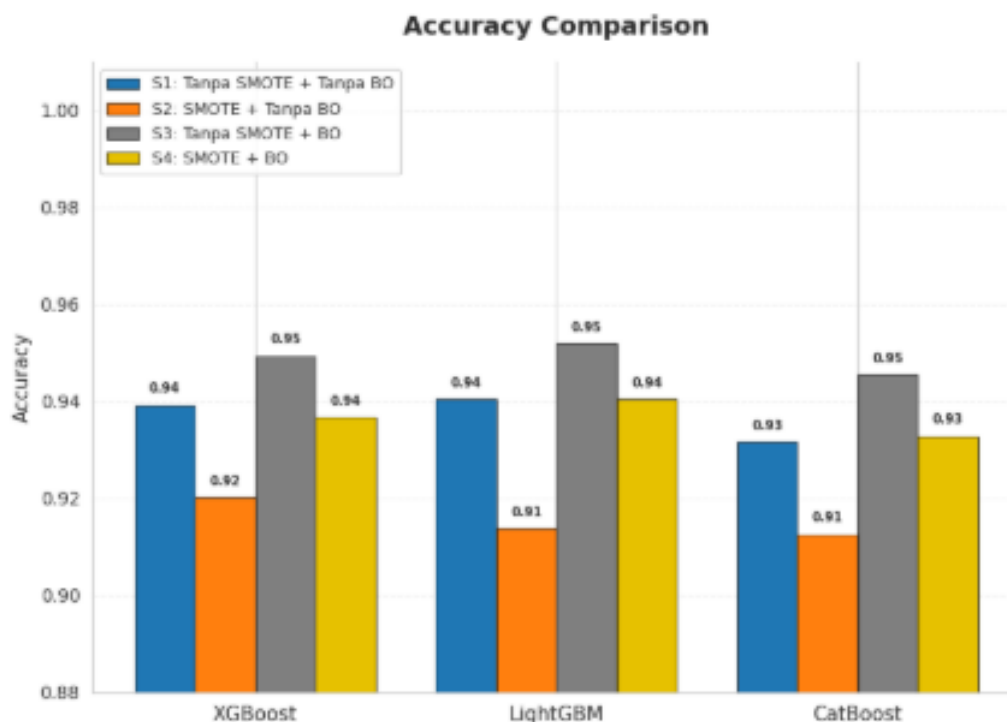
Dari Gambar 6, terlihat bahwa *LightGBM* menghasilkan kurva *ROC* yang paling mendekati sudut kiri atas, yang merepresentasikan *AUC-ROC* tertinggi sebesar 0.9926. Hal ini menunjukkan bahwa *LightGBM* memiliki kemampuan diskriminasi yang superior dibandingkan *XGBoost* (*AUC-ROC* 0.9899) dan *CatBoost* (*AUC-ROC* 0.9885) dalam membedakan pelanggan yang akan churn dan yang bertahan.

Keunggulan *LightGBM* ini dapat dijelaskan oleh arsitektur *histogram-based* dan *leaf-wise tree growth* yang dimilikinya, yang memungkinkan model untuk belajar lebih efisien dari data yang telah di-*resampling* menggunakan *SMOTENC*. Selain itu, kemampuan *LightGBM* dalam menangani fitur kategorikal secara native juga memberikan keuntungan pada dataset *e-commerce* yang memiliki beberapa fitur kategorikal seperti *PreferredOrderCat* dan *MaritalStatus*.

3.6 Ringkasan Performa Terbaik Seluruh Model

Tabel 5. Ringkasan Skenario dan Performa Terbaik Setiap Model

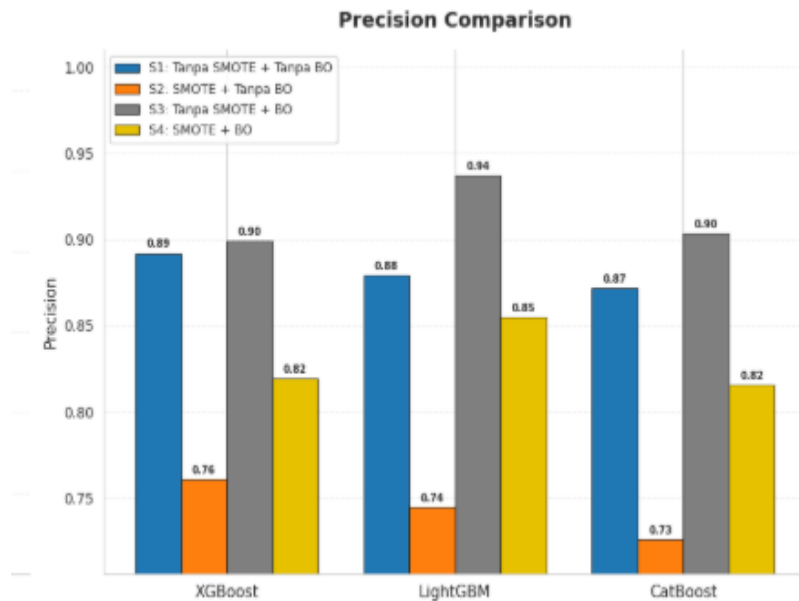
Model	Skenario Terbaik	Accuracy	F1-Score	AUC-ROC
XGBoost	S4: SMOTENC + BO	0.9571	0.9576	0.9899
LightGBM	S4: SMOTENC + BO	0.9633	0.9635	0.9926
CatBoost	S4: SMOTENC + BO	0.9546	0.9550	0.9885



Gambar 8. Nilai Accuracy Setiap Model Algoritma Berdasarkan Skenario S1 – S4

Gambar 8 merupakan *Accuracy Comparison*, grafik ini menunjukkan nilai *Accuracy* (Akurasi) dari ketiga model pada keempat skenario. Tren umum secara visual, skenario S3 (Abu-abu: Tanpa SMOTE + BO) menunjukkan nilai akurasi tertinggi di ketiga model (sekitar 0.95). Skenario S2 (Oranye: SMOTE saja) menunjukkan nilai akurasi terendah (sekitar 0.91 - 0.92). Skenario S1 (Biru: Baseline) dan S4 (Kuning: SMOTE + BO) berada di tengah-tengah (sekitar 0.93 - 0.94).

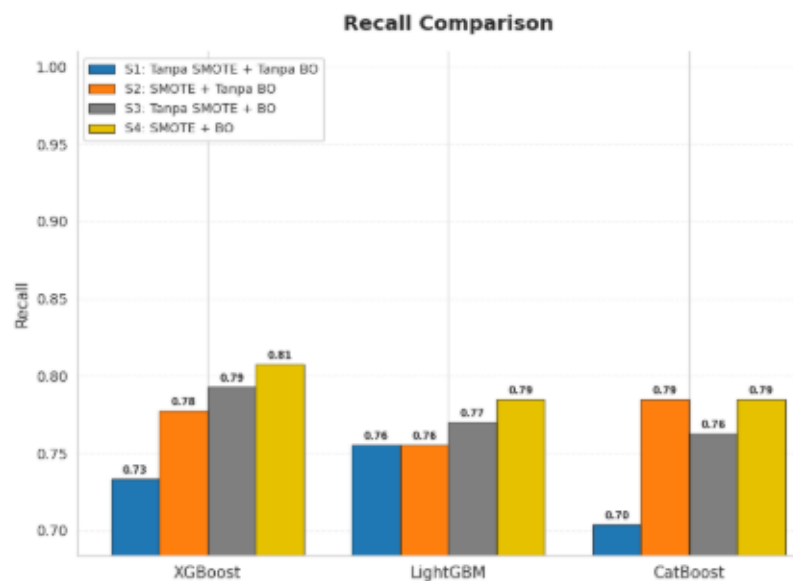
Analisis pengaruh *Bayesian Optimization* (BO), skenario S3 (hanya BO) memiliki akurasi tertinggi. Ini menunjukkan bahwa tuning *hyperparameter* sangat efektif dalam meningkatkan akurasi model secara keseluruhan. Pengaruh SMOTE: Skenario S2 (hanya SMOTE) memiliki akurasi terendah. Hal ini sering terjadi karena SMOTE menciptakan data sintetik untuk kelas minoritas (Churn), yang terkadang dapat memperkenalkan noise atau menyebabkan model overfitting pada kelas minoritas, sehingga akurasi keseluruhan (yang didominasi kelas mayoritas No Churn) bisa sedikit menurun atau tidak meningkat signifikan dibandingkan baseline.



Gambar 9. Nilai Precision Setiap Model Algoritma Berdasarkan Skenario S1 – S4

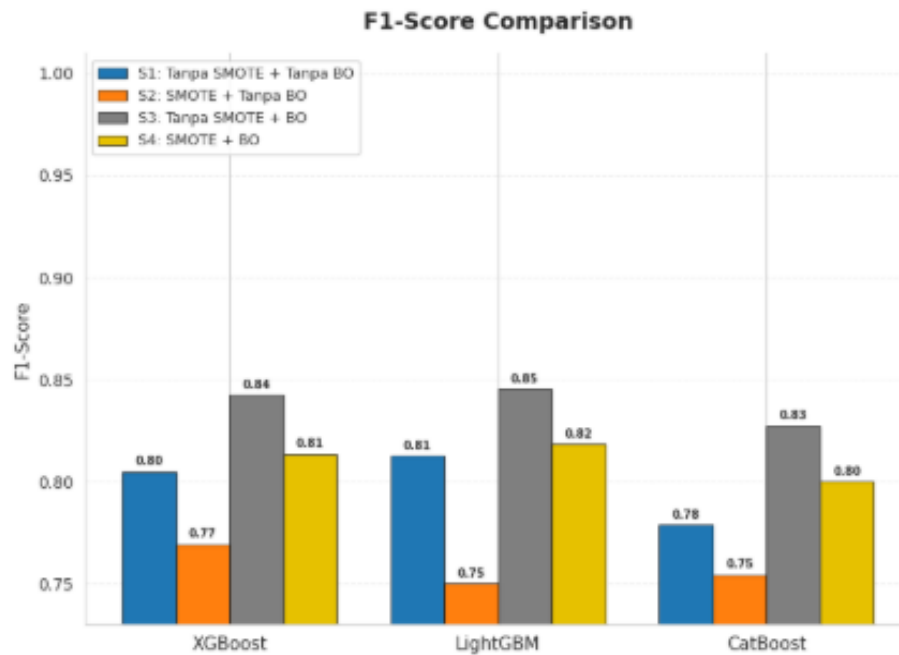
Gambar 9, *Precision Comparison*, grafik ini menunjukkan nilai *Precision* (Presisi) dari ketiga model. *Precision* mengukur seberapa banyak prediksi *Churn* yang benar-benar *Churn* (minim False Positive). Tren umum pola yang sama terlihat di sini. Skenario S3 (Abu-abu: Tanpa SMOTE + BO) memiliki presisi tertinggi (0.90 - 0.94). Skenario S2 (Oranye: SMOTE saja) memiliki presisi terendah secara signifikan (0.73 - 0.76).

Analisis, penurunan Presisi pada S2: Penggunaan SMOTE saja (S2) menurunkan presisi secara drastis. Ini logis karena SMOTE menambah sampel minoritas, yang bisa membuat model lebih "agresif" memprediksi *Churn*, sehingga meningkatkan False Positive (memrediksi *Churn* padahal sebenarnya No *Churn*), yang menurunkan nilai Presisi. Peran *Bayesian Optimization*, Skenario S3 dan S4 memiliki presisi yang lebih baik daripada S2. Ini menunjukkan bahwa optimasi hyperparameter membantu model menyeimbangkan batasan keputusan (decision boundary) sehingga tidak terlalu agresif memprediksi kelas minoritas, menjaga presisi tetap tinggi.



Gambar 10. Nilai Recall Untuk Setiap Model Algoritma Berdasarkan Skenario S1 – S4

Gambar 10 perbandingan *Recall*, grafik pertama menampilkan perbandingan nilai *Recall* untuk tiga algoritma *machine learning* (*XGBoost*, *LightGBM*, dan *CatBoost*) yang diuji dalam empat skenario berbeda. Secara umum, terlihat bahwa penerapan teknik SMOTE (Skenario 2 dan 4) cenderung meningkatkan nilai *Recall* dibandingkan skenario tanpa SMOTE (Skenario 1 dan 3), terutama pada algoritma *XGBoost* dan *CatBoost*. Pada *XGBoost*, terjadi peningkatan yang konsisten dari Skenario 1 (0.73) hingga mencapai nilai tertinggi pada Skenario 4 (0.81). Pada *CatBoost*, lonjakan signifikan terjadi pada Skenario 2 (0.78) dan bertahan tinggi di Skenario 4 (0.78) dibandingkan Skenario 1 (0.70). Hal ini mengindikasikan bahwa teknik SMOTE sangat efektif dalam membantu model mengenali kelas minoritas (pelanggan *Churn*) dengan lebih baik, sehingga mengurangi jumlah False Negative (pelanggan yang diprediksi tidak *Churn* padahal sebenarnya *Churn*).



Gambar 11. Nilai F1-Score Untuk Setiap Model Algoritma Berdasarkan Skenario S1 – S4

Gambar 11 perbandingan *F1-Score*, grafik kedua menunjukkan perbandingan nilai *F1-Score*, yang merupakan rata-rata harmonik dari *Precision* dan *Recall*. Menariknya, tren yang terlihat berbeda dengan *Recall*, di mana Skenario 3 (Tanpa SMOTE + BO) justru menghasilkan nilai *F1-Score* tertinggi untuk ketiga algoritma (*XGBoost*: 0.84, *LightGBM*: 0.85, *CatBoost*: 0.83). Hal ini mengindikasikan bahwa *Bayesian Optimization* (BO) saja mampu menyeimbangkan *Precision* dan *Recall* dengan lebih baik dibandingkan penggunaan SMOTE saja. Sebaliknya, penggunaan SMOTE saja (Skenario 2) justru menghasilkan nilai *F1-Score* yang cenderung lebih rendah atau bahkan terendah (misalnya pada *LightGBM* turun dari 0.81 di S1 menjadi 0.75 di S2). Hal ini mengindikasikan bahwa meskipun SMOTE berhasil meningkatkan *Recall*, ia kemungkinan besar menurunkan *Precision* secara signifikan (meningkatkan *False Positive*), sehingga menurunkan skor F1 secara keseluruhan. Kombinasi SMOTE dan BO (Skenario 4) berhasil memperbaiki kembali *F1-Score* dibandingkan Skenario 2, namun pada grafik ini belum mampu melampaui performa Skenario 3.

Tabel 4 menunjukkan bahwa seluruh algoritma mencapai performa terbaik pada Skenario 4 (SMOTENC + *Bayesian Optimization*), yang mengkonfirmasi efektivitas kombinasi teknik penanganan ketidakseimbangan kelas dan optimasi hyperparameter [14], [15]. *LightGBM* menjadi model teratas dengan *F1-Score* 0.9635 dan *AUC-ROC* 0.9926, diikuti oleh *XGBoost* dan *CatBoost* dengan selisih performa yang relatif kecil.

3.7 Pembahasan

Hasil penelitian ini memiliki implikasi penting bagi bisnis e-commerce dalam upaya mengurangi tingkat customer churn. Pertama, model prediksi churn yang dikembangkan dapat diintegrasikan ke dalam sistem *Customer Relationship Management* (CRM) untuk memberikan peringatan dini (*early warning system*) tentang pelanggan yang berpotensi churn. Dengan *Accuracy* di atas 96% dan *Recall* yang tinggi, sistem dapat mengidentifikasi pelanggan berisiko dengan akurasi yang dapat diandalkan.

Kedua, analisis *feature importance* memberikan wawasan strategis untuk merancang intervensi yang tepat sasaran. Misalnya, pelanggan baru (*low tenure*) yang mengajukan komplain harus menjadi prioritas untuk mendapatkan perhatian khusus dari tim layanan pelanggan. Program *re-engagement* dapat diluncurkan untuk pelanggan yang sudah

lama tidak bertransaksi.

Ketiga, implementasi *LightGBM* yang efisien secara komputasi memungkinkan *deployment* pada sistem produksi dengan sumber daya terbatas, sehingga cocok untuk *e-commerce* skala kecil hingga menengah yang ingin mengadopsi teknologi *machine learning* tanpa investasi infrastruktur yang besar.

Keempat, strategi retensi dapat dipersonalisasi berdasarkan profil risiko *churn* pelanggan. Pelanggan dengan risiko tinggi dapat menerima insentif khusus seperti *cashback* tambahan, diskon eksklusif, atau layanan prioritas.

4. KESIMPULAN

Penelitian ini berhasil mengevaluasi dan membandingkan kinerja tiga algoritma *gradient boosting*, yaitu *XGBoost*, *LightGBM*, dan *CatBoost*, dalam memprediksi *customer churn* pada dataset *E-commerce* melalui empat skenario eksperimen yang berbeda. Berdasarkan hasil evaluasi menggunakan metrik *Accuracy* dan *AUC-ROC*, dapat disimpulkan bahwa algoritma *LightGBM* merupakan model yang paling superior dan konsisten dalam menangani kasus prediksi *churn* ini. Secara keseluruhan, *LightGBM* pada Skenario 3, yang menerapkan *Bayesian Optimization* tanpa *SMOTENC*, mencatatkan performa tertinggi dengan nilai *Accuracy* sebesar 0.9518 dan *AUC-ROC* sebesar 0.9659. Hal ini mengindikasikan bahwa *LightGBM* memiliki kemampuan diskriminasi yang sangat baik serta efisiensi komputasi yang unggul dibandingkan *XGBoost* dan *CatBoost*, bahkan tanpa memerlukan teknik resampling data. Selanjutnya, eksperimen ini membuktikan bahwa penerapan *Bayesian Optimization* menggunakan *framework Optuna* memberikan dampak positif yang signifikan terhadap peningkatan performa model dibandingkan dengan penggunaan parameter default. Di sisi lain, penerapan teknik *SMOTENC* terbukti sangat efektif dalam meningkatkan nilai *Recall* pada kelas minoritas, yang berarti model menjadi jauh lebih sensitif dalam mendeteksi pelanggan yang berpotensi melakukan *churn*. Namun, peningkatan *Recall* ini diiringi dengan fenomena *trade-off* berupa penurunan *Accuracy* keseluruhan. Oleh karena itu, kombinasi *SMOTENC* dan *Bayesian Optimization* pada Skenario 4 lebih direkomendasikan untuk diimplementasikan jika tujuan bisnis utama perusahaan adalah memaksimalkan deteksi dini pelanggan *churn* (meminimalkan false negative) dengan bersedia mengorbankan sedikit akurasi global. Selain evaluasi performa model, analisis *feature importance* mengungkapkan bahwa variabel *Tenure* (lama berlangganan), *Complain* (adanya komplain), *CashbackAmount* (jumlah *cashback*), *DaySinceLastOrder*, dan *MaritalStatus* merupakan prediktor paling dominan yang mempengaruhi keputusan *churn* pelanggan. Temuan ini memberikan implikasi praktis bagi manajemen *e-commerce* untuk merancang strategi retensi yang lebih terarah dan efisien, seperti memberikan perhatian khusus serta intervensi promo pada pelanggan baru dengan masa berlangganan pendek yang pernah mengajukan komplain. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi teknik penanganan ketidakseimbangan kelas alternatif seperti *ADASYN*, mengintegrasikan data perilaku pelanggan yang lebih kaya seperti riwayat browsing, serta melakukan validasi model menggunakan data riil dari platform *e-commerce* di Indonesia untuk mengukur tingkat generalisasi model di konteks praktis.

REFERENCES

- [1] A. Ninov Royana, Y. Vita Via, C. Aji Putra Informatika, U. Pembangunan Nasional, and J. Timur Jalan Raya Rungkut Madya, "EVALUASI KINERJA LIGHTGBM DAN CATBOOST UNTUK PREDIKSI CHURN BERDASARKAN DATASET PELANGGAN LAYANAN STREAMING MUSIK," 2025. [Online]. Available: <https://www.kaggle.com/c/streaming-subscription>
- [2] S. H. Sholeha, M. Faid, and Moh. A. Yaqin, "Prediksi Prediksi Perpindahan Pelanggan Pada Toko Online Menggunakan Metode Tree-Based Gradient Boosted Models," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 3, pp. 605–614, May 2024, doi: 10.47065/josyc.v5i3.5215.
- [3] M. Sakmar, N. T. Kadir, P. A. Shofa, and A. Darmawan, "EFEKTIVITAS XGBOOST LIGHTGBM DAN CATBOOST PADA DATASET IMBALANCED PREDICTIVE MAINTENANCE," *SINTA Jurnal SistemInformasi dan Teknologi Komputasi*, vol. 3, pp. 36–44, 2026, doi: 10.61124/sinta.v3i1.145.
- [4] P. Ridwan Yulianto and Y. Parti Astuti, "Perbandingan Kinerja Algoritma CatBoost, XGBoost, LightGBM dan Random Forest Dalam Memprediksi Risiko Infeksi Aids Dalam Dataset Kesehatan," *Technology and Science (BITS)*, vol. 7, no. 4, 2026, doi: 10.47065/bits.v7i4.8975.
- [5] "Teknik_Weighting_untuk_Mengatasi_Ketidakeimbangan".
- [6] D. Tribuana, B. Baharuddin, and A. Muhammad Resky, "Penerapan Algoritma XGBoost Untuk Prediksi Kepuasan Pelanggan Pada Layanan E-Commerce: Studi Pada Dataset Transaksi Nyata," *Jurnal Teknologi dan Bisnis Cerdas*, vol. 1, no. 1, pp. 50–59, Jul. 2025, doi: 10.64476/jtbc.v1i1.5.
- [7] F. Pratama, E. Ali, Rahmaddeni, and W. Agustin, "Perbandingan Kinerja Xgboost Dan Lightgbm Dalam Klasifikasi Depresi Pada Mahasiswa Berdasarkan Faktor Demografi Dan Akademik," *Jurnal Algoritma*, vol. 22, no. 2, pp. 53–64, Nov. 2025, doi: 10.33364/algoritma/v.22-2.2439.
- [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] R. Fadly, “Perbandingan Algoritma Catboost dan Extra Trees Classifier Untuk Prediksi Tingkat Keberhasilan Panen Padi Berdasarkan Faktor Cuaca dan Kondisi Tanah,” 2026. [Online]. Available: <https://ejournal.edutechjaya.com/index.php/komprehensif>
- [10] C. Valentino *et al.*, “Analisis Kinerja XGBoost Menggunakan Bayesian Optimization dalam Prediksi Harga Ethereum,” *JNATIA*, vol. 3, no. 4, 2025.
- [11] F. Nuria Handayani, V. Atina, and A. Arumsari, “BULLETIN OF COMPUTER SCIENCE RESEARCH Penerapan XGBoost dan SMOTE untuk Klasifikasi Metode Pembayaran Pelanggan pada Data Transaksi Tidak Seimbang,” *Media Online*, vol. 6, no. 4, pp. 1498–1508, 2026, doi: 10.47065/bulletincsr.v6i4.1127.
- [12] T. A. E. Putri, T. Widiarini, and R. Santoso, “PENERAPAN TUNING HYPERPARAMETER RANDOMSEARCHCV PADA ADAPTIVE BOOSTING UNTUK PREDIKSI KELANGSUNGAN HIDUP PASIEN GAGAL JANTUNG,” *Jurnal Gaussian*, vol. 11, no. 3, pp. 397–406, Jan. 2023, doi: 10.14710/j.gauss.11.3.397-406.
- [13] D. Nuraini, D. Violina, D. R. Anamisa, B. K. Khotimah, A. Jauhari, and F. A. Mufarroha, “Prediksi Hasil Panen Padi dengan Metode Multiple Linear Regression dan Particle Swarm Optimization untuk Meningkatkan Produksi Padi di Madura,” *JUSIFOR : Jurnal Sistem Informasi dan Informatika*, vol. 4, no. 1, pp. 1–8, Jun. 2025, doi: 10.70609/jusifor.v4i1.5857.
- [14] Y. Irawan, R. Wahyuni, and R. Ordila, “Comparative Analysis of Machine Learning Algorithms with SMOTE and Boosting Techniques in Accuracy Improvement,” vol. 13, no. 5, 2024, doi: 10.33022/ijcs.v13i5.4386.
- [15] O. Pahlevi, D. A. N. Wulandari, L. K. Rahayu, H. Leidiyana, and Y. Handrianto, “Model Klasifikasi Risiko Stunting Pada Balita Menggunakan Algoritma CatBoost Classifier,” *Bulletin of Computer Science Research*, vol. 4, no. 6, pp. 414–421, Oct. 2024, doi: 10.47065/bulletincsr.v4i6.373.