

Evaluasi Dampak SMOTE-NC Terhadap Stabilitas Interpretasi SHAP pada Model XGBoost Teroptimasi untuk Klasifikasi *Churn* Pelanggan

Haeruzzikri^{*1}, Robert Marco²

^{1,2}PJJ Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Indonesia
Email: ¹haeruzzikri@students.amikom.ac.id, ²robertmarco@amikom.ac.id

Abstrak

Ketidakeimbangan kelas pada dataset telekomunikasi dapat memicu tingginya angka false negative dalam prediksi customer churn. Pendekatan SMOTE-NC mampu mengatasi masalah tersebut, namun berisiko mendistorsi interpretasi pada model black-box seperti XGBoost. Penelitian ini bertujuan mengevaluasi efektivitas SMOTE-NC dalam meningkatkan deteksi kelas minoritas, sekaligus mengukur stabilitas interpretasi Explainable AI (XAI) pada XGBoost yang dioptimasi dengan Optuna melalui evaluasi multi-seed dan skenario kontrol (ablation study). Ekstraksi fitur dievaluasi dengan TreeSHAP, sementara stabilitas kuantitatif diukur melalui Jaccard Index dan Spearman Rank Correlation pada variasi ambang batas $k = 5, 10, 15$. Hasil empiris menunjukkan SMOTE-NC secara konsisten meningkatkan kepekaan deteksi (Recall) sebesar +8.55%, dengan kompromi penurunan akurasi global sebesar -2.38%. Analisis XAI mengonfirmasi bahwa kendati terjadi distorsi interpretasi pada representasi fitur kategorikal, komposisi fitur utama ($k = 10$) secara umum berhasil dipertahankan secara konsisten (rerata Jaccard Index 0.7939 ± 0.1233) dengan korelasi hierarki yang sangat kuat (Spearman Correlation 0.8526). Pengujian kontrol membuktikan bahwa optimasi hyperparameter Optuna berfungsi memitigasi degradasi stabilitas eksplanasi yang dipicu oleh SMOTE-NC. Penelitian ini menyimpulkan bahwa kompromi performa akibat interpolasi data sintesis layak secara operasional untuk strategi retensi pelanggan, dengan prioritas intervensi bisnis pada fitur jenis kontrak, masa berlangganan, dan metode pembayaran.

Kata kunci: Churn Pelanggan, Explainable AI, SHAP, SMOTE-NC, XGBoost.

Abstract

Class imbalance in telecommunication datasets can trigger a high rate of false negatives in customer churn prediction. While the SMOTE-NC approach can address this issue, it risks distorting feature interpretations in black-box models such as XGBoost. This study evaluates SMOTE-NC's effectiveness in improving minority-class detection while measuring the stability of Explainable AI (XAI) interpretations on Optuna-optimized XGBoost through multi-seed evaluation and a control scenario (ablation study). We evaluated feature importance using TreeSHAP, and measured quantitative stability via the Jaccard Index and Spearman Rank Correlation across threshold variations of $k = 5, 10, 15$. Empirical results show that SMOTE-NC consistently improves detection sensitivity (Recall) by +8.55%, with a trade-off of a -2.38% reduction in global accuracy. XAI analysis confirms that despite interpretation distortions in categorical feature representations, the core top-10 feature composition is consistently preserved (mean Jaccard Index of 0.7939 ± 0.1233) with a very strong hierarchical correlation (Spearman Correlation of 0.8526). The control experiment demonstrates that Optuna hyperparameter optimization functions effectively to mitigate the degradation of explanation stability induced by SMOTE-NC. This study concludes that the performance trade-off resulting from synthetic data interpolation is operationally viable for customer retention strategies, with business intervention prioritized on contract type, tenure, and payment method features.

Keywords: Customer Churn, Explainable AI, SHAP, SMOTE-NC, XGBoost.

This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)



1. PENDAHULUAN

Tingkat persaingan yang tinggi antar penyedia layanan *Internet Service Provider* (ISP) memicu peningkatan fenomena *customer churn*, yakni kondisi di mana pelanggan memutuskan untuk berhenti berlangganan dan berpindah ke penyedia kompetitor[1]. Mengingat biaya operasional untuk mengakuisisi pelanggan baru jauh lebih besar dan kurang efisien dibandingkan upaya mempertahankan pelanggan eksisting, kapabilitas untuk memprediksi probabilitas *churn* secara presisi menjadi fondasi

strategis bagi keberlanjutan bisnis telekomunikasi[2],[3]. Meskipun pendekatan analitik saat ini telah bergeser ke *Machine Learning*, data industri telekomunikasi sering kali dihadapkan pada masalah ketidakseimbangan kelas (*class imbalance*). Kondisi ini menyebabkan model cenderung bias terhadap kelas mayoritas (pelanggan yang tidak *churn*)[4]. Bias ini memicu tingginya risiko *false negative* (kegagalan sistem mendeteksi pelanggan yang sebenarnya berpotensi *churn*), yang berdampak langsung pada hilangnya pendapatan perusahaan[5].

Untuk mengatasi ketidakseimbangan tersebut, pendekatan *Synthetic Minority Over-sampling Technique for Nominal and Continuous* (SMOTE-NC) terbukti krusial untuk menyeimbangkan campuran atribut kategorikal dan numerik pada data pelanggan[6]. Namun, teknik injeksi data sintesis ini memiliki kelemahan, yakni berpotensi mendistorsi distribusi data asli beserta logika pengambilan keputusan model[7],[8],[9]. Tantangan ini semakin kompleks saat berhadapan dengan algoritma *ensemble* seperti XGBoost[10],[11]. Akurasi prediktif yang tinggi dari model *black-box* tidak memberikan nilai tambah jika manajemen tidak dapat memahami faktor mengapa pelanggan memilih untuk berhenti[12].

Tinjauan literatur menunjukkan adanya tantangan dalam menyeimbangkan akurasi prediktif dan interpretabilitas pada data tidak seimbang (*class imbalance*). Studi terdahulu hanya berfokus pada salah satu aspek secara terpisah. Penelitian oleh Imani dan Arabia 2023 yang telah mengintegrasikan XGBoost, SMOTE dan Optuna untuk mengoptimalkan klasifikasi *churn* dengan data tidak seimbang pada sektor telekomunikasi[13]. Namun, penelitian tersebut mengevaluasi model berdasarkan metrik performa teknis tanpa melibatkan *Explainable AI* (XAI). Perkembangan terbaru dalam pemodelan prediksi *churn* telekomunikasi mulai mengintegrasikan teknik penyeimbangan data dengan kerangka kerja XAI. Integrasi ini diterapkan oleh penelitian Andy Hermawan dkk. 2024 mengintegrasikan algoritma *Gradient Boosting* dengan memanfaatkan teknik *oversampling* SMOTE konvensional dan ADASYN untuk mengatasi tingginya rasio *false negative* pada metode *rule-based* konvensional, lalu menggunakan SHAP untuk mengungkap pola kompleks perilaku pelanggan yang sebelumnya tidak terdeteksi[14]. Pendekatan serupa diadopsi oleh Asif dkk. 2025 menerapkan teknik SMOTE konvensional pada model *ensemble* TriBoost guna menjembatani kebutuhan akan akurasi prediktif yang tinggi dengan tingkat interpretabilitas model. Melalui kerangka LIME dan SHAP, studi tersebut menjelaskan prediksi individual dan menyoroti fitur-fitur utama pemicu *churn* secara transparan[15], serta El Attar dan El Hajj 2026 yang mengombinasikan *soft-voting ensemble*, *deep autoencoder*, dan SHAP pada data *churn* telekomunikasi[16]. Meskipun studi-studi tersebut telah berhasil melibatkan XAI pada kondisi data yang diseimbangkan, implementasi SHAP pada studi - studi tersebut masih terbatas pada fungsi ekstraksi nilai dan visualisasi statis.

Evaluasi mengenai stabilitas interpretasi XAI akibat teknik penyeimbangan kelas mulai digunakan pada beberapa domain lain. Dalam ranah deteksi penipuan finansial, Alkhozai dkk. 2025 memverifikasi konsistensi nilai SHAP setelah penerapan algoritma ADASYN[17]. Pada domain risiko media, Mammadov dkk. 2026 mengonfirmasi adanya degradasi eksplanasi SHAP akibat ketidakseimbangan kelas melalui pengukuran kuantitatif[18]. Selanjutnya, studi empiris lintas domain oleh Kim 2026 menunjukkan bahwa teknik manipulasi data seperti SMOTE konvensional dapat menurunkan stabilitas eksplanasi hierarki fitur pada struktur SHAP[19]. Implikasi dari temuan Kim ini membuka ruang lingkup penelitian baru pada sektor telekomunikasi. Mengingat karakteristik data pelanggan telekomunikasi umumnya bersifat tabular campuran (*mixed nominal-continuous*), penyeimbangan data memerlukan varian khusus seperti SMOTE-NC. Hingga saat ini, analisis kuantitatif mengenai pengaruh SMOTE-NC terhadap pergeseran peringkat fitur pada struktur SHAP belum dieksplorasi, terutama pada model XGBoost yang dioptimasi dengan Optuna.

Oleh karena itu, penelitian ini bertujuan mengukur *trade-off* antara efektivitas SMOTE-NC dalam menekan *false negative* dan potensi distorsi terhadap stabilitas interpretasi SHAP. Kebaruan penelitian ini berfokus pada dua aspek utama, penerapan SMOTE-NC yang dirancang khusus untuk menangani interaksi data campuran kategorikal dan numerik pada sektor telekomunikasi, berbeda dari studi terdahulu yang menggunakan SMOTE konvensional, serta pengukuran pergeseran matematis pada arsitektur TreeSHAP secara kuantitatif menggunakan metrik *Jaccard Similarity* dan *Spearman Rank Correlation* untuk menguji retensi serta stabilitas hierarki peringkat fitur, bukan sebatas visualisasi deskriptif statis.

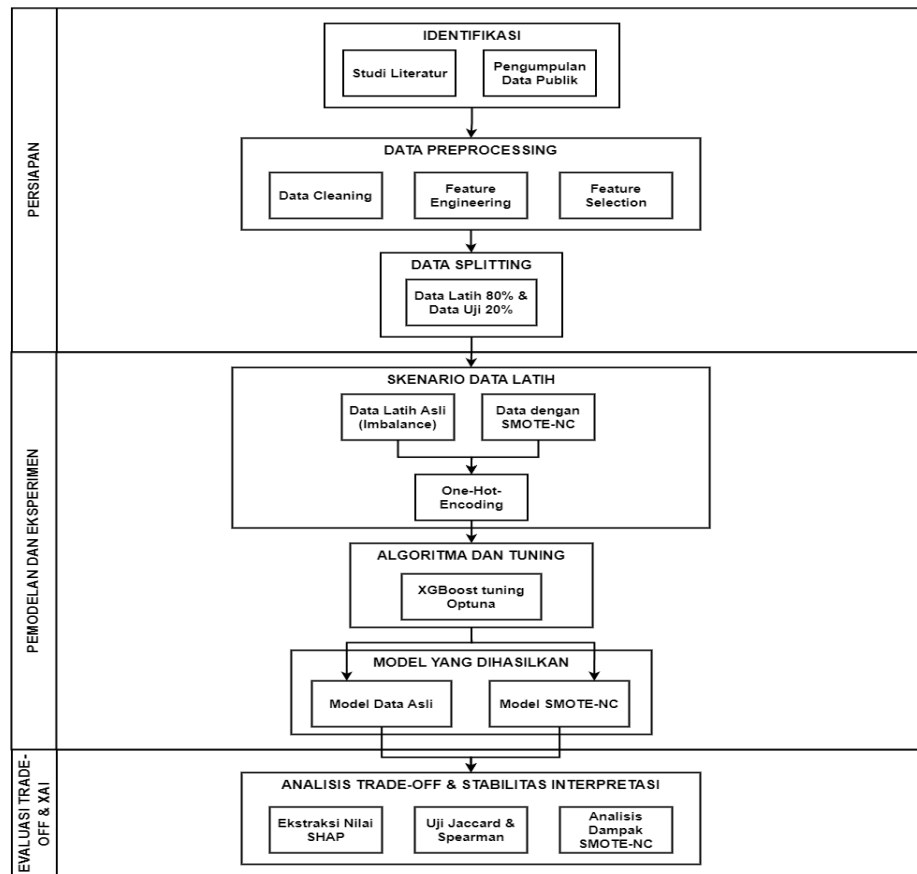
Untuk memberikan arah evaluasi yang terukur, eksperimen dalam penelitian ini merumuskan dua pertanyaan penelitian utama:

1. RQ1 : Bagaimana efektivitas integrasi SMOTE-NC dan Optuna pada model XGBoost dalam mereduksi tingkat *false negative* pada data pelanggan telekomunikasi yang tidak seimbang?
2. RQ2 : Apakah penerapan teknik data sintetis tipe campuran tersebut memicu pergeseran matematis yang memengaruhi stabilitas peringkat fitur pemicu *churn* pada struktur SHAP?

Guna memastikan arah evaluasi yang mendalam dan terfokus, ruang lingkup penelitian ini dibatasi pada beberapa aspek metodologis yang spesifik. Eksperimen dilakukan hanya menggunakan dataset publik *customer churn* penyedia layanan internet (*broadband ISP*) yang memiliki karakteristik struktur data tabular campuran (*mixed nominal-continuous*). Intervensi penanganan ketidakseimbangan kelas dibatasi secara eksklusif pada implementasi algoritma SMOTE-NC tanpa melakukan komparasi dengan varian teknik *oversampling* lainnya. Proses klasifikasi diterapkan menggunakan algoritma XGBoost, di mana kerangka Optuna yang difungsikan sebagai instrumen prasyarat untuk optimasi hiperparameter model sebelum fase interpretasi dilakukan, bukan sebagai objek perbandingan metode pencarian parameter. Selanjutnya, analisis transparansi model dibatasi pada penggunaan kerangka kerja SHAP (*TreeExplainer*) untuk mengevaluasi peringkat fitur penting secara global (*global feature importance*). Terakhir, evaluasi performa model berfokus pada metrik *Recall* (sensitivitas kelas minoritas) untuk mereduksi tingkat *false negative* serta nilai *F1-Score*, sedangkan evaluasi stabilitas eksplanasi XAI diukur secara kuantitatif melalui metrik *Jaccard Similarity* dan *Spearman Rank Correlation*.

2. METODE PENELITIAN

Penelitian ini merupakan eksperimen komputasional kuantitatif yang mengadopsi pendekatan pengujian komparatif. Eksperimen dirancang untuk membandingkan kinerja dan stabilitas pemodelan algoritma Extreme Gradient Boosting (XGBoost) melalui tiga skenario utama: Skenario 1 (Data Asli/Baseline), Skenario 2 (SMOTE-NC dengan Optimasi Optuna), dan Skenario Kontrol (Ablation Study berbasis data SMOTE-NC dengan hyperparameter baseline). Guna menjamin validitas statistik dan mengisolasi bias pengacakan sampel serta faktor perancu hyperparameter, seluruh skenario dieksekusi secara berulang menggunakan 5 random seeds berbeda ($N = 5$). Evaluasi analitis dititikberatkan pada kuantifikasi trade-off antara efektivitas reduksi False Negative dan tingkat distorsi stabilitas fitur pada struktur interpretasi Explainable AI berbasis TreeSHAP.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data dan Karakteristik Data

Penelitian ini menggunakan dataset bersumber dari Kaggle bernama Telco Customer Churn[20]. Pemilihan dataset didasarkan pada karakteristik data yang relevan dengan kebutuhan metodologis. Pertama, dengan total 7.043 rekam data dan 21 atribut heterogen (demografi, layanan, dan finansial), dataset ini memiliki struktur data campuran (*mixed nominal-continuous*). Karakteristik tersebut diperlukan untuk menguji algoritma SMOTE-NC yang dirancang khusus untuk tipe data campuran, yang tidak dapat ditangani secara optimal oleh SMOTE konvensional. Kedua, dataset ini memiliki ketidakseimbangan kelas (*class imbalance*) dengan rasio 73,5% kelas mayoritas *Non-churn* (5.174 data) dan 26,5% kelas minoritas *Churn* (1.869 data). Ketiga, penggunaan dataset yang telah menjadi standar tolok ukur (*benchmark*) dalam literatur prediksi *churn* ini mendukung aspek replikasi dan komparabilitas hasil analisis stabilitas interpretasi SHAP secara valid dengan penelitian-penelitian terdahulu.

Tabel 1. Karakteristik Dataset

No	Nama Atribut	Tipe Data Awal
1	customerID	Object
2	gender	Object
3	SeniorCitizen	Int64
4	Partner	Object
5	Dependents	Object
6	tenure	Int64
7	PhoneService	Object
8	MultipleLines	Object
9	InternetService	Object
10	OnlineSecurity	Object
11	OnlineBackup	Object

No	Nama Atribut	Tipe Data Awal
12	DeviceProtection	Object
13	TechSupport	Object
14	StreamingTV	Object
15	StreamingMovies	Object
16	Contract	Object
17	PaperlessBilling	Object
18	PaymentMethod	Object
19	MonthlyCharges	Float64
20	TotalCharges	Object*
21	Churn	Object

2.2 Analisis dan Pemodelan Data

Alur eksperimen komputasi dijalankan melalui serangkaian tahapan *pipeline machine learning* yang terisolasi ketat untuk menjaga kemurnian data serta menjamin keterulangan (*reproducibility*) hasil eksperimen. Seluruh eksperimen dieksekusi pada lingkungan Python 3.12 dengan spesifikasi versi pustaka utama meliputi imbalanced-learn 0.14.2, scikit-learn 1.6.1, xgboost 3.2.0, optuna 4.9.0, shap 0.52.0, dan scipy 1.16.3.

2.2.1 Prapemrosesan Data (Preprocessing)

Meliputi konversi tipe data paksa (*coercion*) pada fitur *TotalCharges* menjadi *float64* dengan menghapus 11 baris *missing values* secara *listwise deletion* guna mencegah bias imputasi. Selanjutnya, 18 variabel kategorikal sengaja dipertahankan dalam format teks aslinya (*string/object*) guna menjaga keabsahan identitas spasial sebelum sintesis data. Tahap ini diakhiri dengan penghapusan fitur administratif *customerID* (reduksi dimensi) untuk mengeliminasi *noise* komputasi. Transformasi binerisasi (*One-Hot Encoding*) ditunda dan baru dieksekusi pasca-sintesis data tepat sebelum pemodelan.

2.2.2 Pemisahan Data (Data Splitting)

Dataset bersih dipisahkan menjadi data latih (*training set*, 80% atau 5.625 baris) dan data uji (*testing set*, 20% atau 1.407 baris) menggunakan fungsi *train_test_split* secara terstratifikasi (*stratify=y*). Guna mengisolasi bias variansi acak pada pembagian data tunggal serta memperkuat validitas hasil eksperimen, proses pemisahan data dieksekusi secara berulang menggunakan 5 *random seeds* berbeda (*random_state* ∈ 42, 100, 2024, 777, 888).

2.2.3 Penanganan Ketidakseimbangan Kelas

Algoritma *Synthetic Minority Over-sampling Technique for Nominal and Continuous* (SMOTE-NC) diinjeksikan secara eksklusif hanya pada data latih dengan konfigurasi parameter *k_neighbors* = 5, *categorical_features=cat_indices*, dan *random_state=42*. Penanganan atribut kategorikal pada SMOTE-NC dilakukan sebelum tahapan *One-Hot Encoding*. Jika *One-Hot Encoding* diterapkan sebelum SMOTE-NC, variabel biner hasil enkoding akan diperlakukan sebagai data kontinu yang menghasilkan nilai sintesis decimal yang tidak logis. Perhitungan jarak antara titik observasi kelas minoritas *x* dan kandidat tetangganya *y* pada data campuran dihitung menggunakan jarak Euclidean termodifikasi dengan penalti deviasi standar sesuai persamaan berikut:

$$d(x, y) = \sqrt{\sum_{i \in \text{kontinu}} (x_i - y_i)^2 + m \cdot \sigma_{med}^2} \quad (1)$$

Dimana:

- $d(x, y)$: Jarak Euclidean termodifikasi antara sampel data x dan sampel kandidat y pada ruang data campuran (*mixed space*).
- x : Titik data observasi dari kelas minoritas.
- y : Titik data dari salah satu tetangga terdekat (*k*-nearest neighbor) pada kelas minoritas.

- i : Indeks untuk atribut atau fitur numerik (kontinu).
- x_i, y_i : Nilai atribut kontinu ke- i masing-masing pada sampel x dan y .
- m : Jumlah total atribut kategorikal (nominal) yang memiliki nilai tidak sama antara sampel x dan y ($x_j \neq y_j$).
- σ_{med} : Nilai median deviasi standar dari seluruh fitur kontinu pada kelas minoritas.
- σ_{med}^2 : Faktor penalti kuadrat deviasi standar median yang ditambahkan untuk setiap ketidakcocokan nilai pada atribut kategorikal.

Setelah k -tetangga terdekat diisolasi, sampel sintetis baru ($\tilde{x}$) dibentuk menggunakan persamaan:

$$\tilde{x}_i = x_i + \lambda \cdot (y_i - x_i), \quad \lambda \sim U(0,1) \quad (2)$$

Dimana:

- $\tilde{x}_i$: Nilai atribut kontinu ke- i untuk sampel sintetis baru ($\tilde{x}$) yang dihasilkan.
- x_i : Nilai atribut kontinu ke- i dari sampel acuan kelas minoritas x .
- y_i : Nilai atribut kontinu ke- i dari tetangga terdekat terpilih y .
- λ : Bilangan acak (*random weight*) yang ditarik dari distribusi seragam kontinu dalam rentang 0 hingga 1.
- $U(0,1)$: Notasi untuk distribusi seragam (*uniform distribution*) pada rentang [0,1].

Nilai atribut kategorikal untuk sampel buatan $\tilde{x}$ dibentuk berdasarkan nilai mayoritas (*majority vote* atau modus) dari k -tetangga terdekatnya. Pendekatan ini memastikan bahwa atribut nominal pada data buatan tetap berbentuk nilai diskrit asli yang valid seperti "Ya" atau "Tidak". Setelah data latih seimbang, seluruh fitur kategorikal ditransformasikan menggunakan *One-Hot Encoding*.

Guna mencegah kebocoran data (*data leakage*), sintesis SMOTE-NC dan *One-Hot Encoding* diisolasi ketat di dalam skema *Stratified 5-Fold Cross-Validation* pada data latih. Proses sintesis data dan pembentukan struktur encoding dieksekusi secara eksklusif hanya pada *Training Fold* (80%), sementara *Validation Fold* (20%) dan Data Uji akhir (20%) disesuaikan (*aligned*) mengikuti struktur matriks kolom data latih mengisi kategori yang tidak muncul dengan angka 0 tanpa pernah terpapar oleh proses pembuatan data sintetis.

2.2.4 Optimasi Model dengan Optuna

XGBoost diterapkan sebagai algoritma utama pengklasifikasi dengan mengunci kriteria evaluasi internal *eval_metric='logloss'* dan *random_state=42*. Guna memastikan struktur pohon keputusan (*decision tree*) terkalibrasi maksimal sebelum dibedah logika interpretasinya, dilakukan pencarian *hyperparameter* dinamis menggunakan kerangka Optuna sebanyak 50 *trials* dengan kriteria optimasi memaksimalkan skor F1-Score (*direction='maximize'*) pada *Stratified 5-Fold Cross-Validation* (*shuffle=True, random_state=42*).

Pencarian berbasis *Tree-structured Parzen Estimator* (TPE) dengan *seed* terkunci (TPESampler(*seed=42*)) mengeksplorasi ruang pencarian terikat[16],[17]. Rincian konfigurasi rentang ruang pencarian *hyperparameter* disajikan pada Tabel 2.

Tabel 2. Ruang Pencarian Hyperparameter Optuna

Nama Hyperparameter	Tipe Data	Rentang Pencarian	Fungsi Kerangka Optuna
learning_rate	Float	0.01 – 0.3	suggest_float(log=True)
max_depth	Integer	3 – 10	suggest_int
n_estimators	Integer	100 – 1000	suggest_int
subsample	Float	0.5 – 1.0	suggest_float
colsample_bytree	Float	0.5 – 1.0	suggest_float
gamma	Float	0 – 5	suggest_float

Pencarian *hyperparameter* dieksekusi secara independen pada data *Baseline* dan *SMOTE-NC* untuk membandingkan performa puncak masing-masing skenario secara adil, sebab perubahan distribusi data akibat SMOTE-NC secara alami memerlukan arsitektur pohon (*max_depth* dan *n_estimators*) yang berbeda. Sebagai pelengkap, skenario kontrol (*ablation study*) dengan

hyperparameter identik disisipkan untuk mengisolasi variabel pengganggu, membuktikan secara empiris apakah pergeseran interpretasi SHAP murni disebabkan oleh intervensi SMOTE-NC atau akibat variasi kedalaman pohon.

2.3 Evaluasi Performa dan Stabilitas SHAP

Tahap akhir melibatkan ekstraksi dan komparasi hasil model menggunakan dua dimensi evaluasi.

2.3.1 Evaluasi Trade-off Performa

Evaluasi performa model klasifikasi dilakukan untuk mengukur efektivitas pereduksian *False Negative* pada kelas minoritas (*churn*) sekaligus menguji keseimbangan prediksi secara menyeluruh[23]. Metrik evaluasi dihitung berdasarkan *Confusion Matrix* melalui persamaan-persamaan berikut:

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

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

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{ROC-AUC} = \int_0^1 \text{TPR}(f) d(\text{FPR}(f)) \quad (7)$$

Di mana:

- TP (True Positive) : Pelanggan *churn* yang berhasil dideteksi dengan benar oleh model.
- TN (True Negative) : Pelanggan *non-churn* yang berhasil diprediksi dengan benar
- FN (False Negative) : Pelanggan *churn* yang gagal dideteksi.
- FP (False Positive) : Pelanggan setia (*non-churn*) yang keliru diprediksi sebagai *churn*.
- $\text{TPR}(f)$: True Positive Rate ($\frac{TP}{TP+FN}$) pada nilai ambang batas (*threshold*) f .
- $\text{FPR}(f)$: False Positive Rate ($\frac{FP}{TN+FP}$) pada nilai ambang batas (*threshold*) f .

Guna menyajikan analisis ketidakpastian yang objektif dari 5 pengujian *random seed* ($N = 5$), seluruh metrik performa prediktif disajikan dalam bentuk nilai rata-rata (μ) dan simpangan baku (σ) yang dihitung menggunakan persamaan berikut:

$$\mu = \frac{1}{N} \sum_{i=1}^N M_i \quad (8)$$

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (M_i - \mu)^2} \quad (9)$$

Di mana:

- $N = 5$: menyatakan jumlah *random seed* pengujian
- M_i : menyatakan nilai metrik evaluasi pada iterasi *seed* ke- i .

2.3.2 Evaluasi Stabilitas SHAP

Tahap ini merupakan inti analisis dari penelitian untuk mengkuantifikasi dampak manipulasi data sintesis oleh SMOTE-NC terhadap transparansi logika model. Evaluasi dilakukan dengan mengekstraksi nilai kontribusi fitur secara global menggunakan kerangka kerja *Explainable AI*,

khususnya algoritma TreeSHAP (*Tree Shapley Additive exPlanations*). Proses ekstraksi nilai SHAP dieksekusi secara eksklusif menggunakan data uji (*testing set*) yang murni dan terisolasi.

Kalkulasi nilai SHAP global untuk setiap fitur ditentukan berdasarkan rata-rata absolut dari nilai SHAP individual di seluruh sampel data uji dengan persamaan berikut:

$$I_j = \frac{1}{M} \sum_{i=1}^M |\phi_j^{(i)}| \quad (10)$$

Di mana:

- I_j : nilai kepentingan global (*global importance*) untuk fitur j
- M : merepresentasikan total jumlah rekam data pada data uji
- $\phi_j^{(i)}$: nilai SHAP (kontribusi marginal) dari fitur j pada sampel data ke- i .

Proses ekstraksi ini menghasilkan dua set daftar peringkat kepentingan fitur (*feature importance ranking*), yakni satu daftar dari model dengan data tidak seimbang dan satu daftar dari model dengan data hasil SMOTE-NC. Daftar peringkat tersebut dibatasi pada 10 fitur teratas (*Top-10*). Pemilihan batasan *Top-10* ini didasarkan pada tiga poin pertimbangan metodologis sebagai berikut:

- Meminimalkan bias pengukuran akibat *noise* pada fitur-fitur terbawah yang nilai SHAP-nya tidak signifikan.
- Cakupan 10 fitur lebih memadai dibandingkan *Top-5* untuk menangkap dinamika pergeseran hierarki fitur yang masih memiliki bobot prediktif.
- Batasan 10 fitur mempertimbangkan kapasitas kognitif manusia (*human-centric*) agar hasil analisis tetap relevan untuk ditindaklanjuti oleh tim bisnis.

Perubahan stabilitas dan potensi distorsi di antara kedua daftar peringkat *Top-10* tersebut kemudian diukur secara matematis menggunakan metrik *Jaccard Similarity* dan *Spearman Rank Correlation*. Sebagai bentuk uji ketahanan (*robustness check*), evaluasi stabilitas ini juga diuji pada variasi ambang batas parameter k yang berbeda, yaitu $k = 5, 10, 15$.

2.3.2.1 Jaccard Similarity Index

Jaccard Similarity (Indeks Jaccard) merupakan ukuran metrik statistik yang digunakan untuk menilai derajat kemiripan atau ketakmiripan antara dua himpunan objek data[24]. Metrik ini diimplementasikan untuk mengukur konsistensi komposisi elemen dengan menguji retensi atau hilangnya fitur-fitur krusial dari kelompok peringkat teratas (*top-k features*) setelah data latih diintervensi oleh SMOTE-NC. Persamaan Jaccard Similarity didefinisikan sebagai berikut:

$$J(A_k, B_k) = \frac{|A_k \cap B_k|}{|A_k \cup B_k|} \quad (11)$$

Di mana:

- A_k : himpunan bagian (subset) fitur yang masuk dalam peringkat k teratas (top-k) pada model basis (baseline).
- B_k : himpunan bagian (subset) fitur yang masuk dalam peringkat k teratas (top-k) pada model intervensi (SMOTE-NC).
- $\cap$: jumlah fitur yang sama-sama muncul di kelompok top-k pada kedua model.
- $\cup$: total variasi fitur unik yang muncul di kelompok top-k dari kedua model.

Nilai indeks Jaccard berkisar antara 0 dan 1. Jika nilai mendekati 1, hal ini menunjukkan bahwa komposisi fitur-fitur paling berpengaruh sebagai pemicu churn tetap stabil dan mampu bertahan meskipun data telah dimanipulasi secara sintesis.

2.3.2.2 Spearman Rank Correlation

Jaccard Similarity hanya berfokus pada ada atau tidaknya suatu fitur di kelompok teratas tanpa memedulikan urutannya, Spearman Rank Correlation digunakan khusus untuk mengukur tingkat pergeseran peringkat (*rank displacement*) secara linier. Metrik ini mengevaluasi apakah fitur yang tadinya diidentifikasi paling krusial oleh model basis mengalami degradasi atau lompatan urutan posisi yang signifikan akibat efek samping SMOTE-NC[19],[20]. Koefisien korelasi Spearman dihitung menggunakan persamaan:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (12)$$

Di mana:

- ρ : koefisien korelasi peringkat Spearman antara kedua model.
- d_i : selisih posisi peringkat (difference in rank) untuk fitur ke- i antara daftar model basis dan model intervensi ($Rank(A_i) - Rank(B_i)$).
- n : total jumlah fitur keseluruhan yang dievaluasi di dalam model.

Nilai ρ berkisar antara -1 hingga 1. Nilai korelasi yang mendekati 1 mengindikasikan bahwa intervensi SMOTE-NC memiliki tingkat stabilitas yang tinggi karena tidak merusak tata urutan logis dari fitur-fitur pemicu customer churn saat dibedah oleh SHAP.

3. HASIL DAN PEMBAHASAN

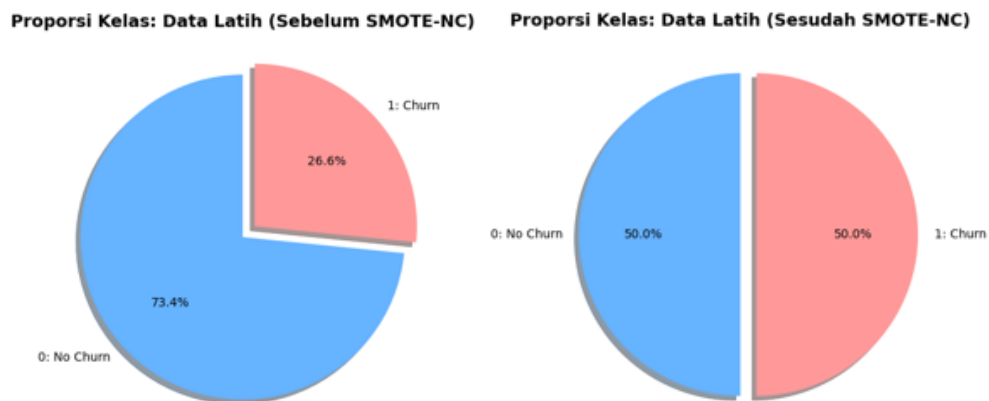
3.1 Prapemrosesan dan Skenario Pemodelan

Dataset *Telco Customer Churn* dengan 7032 observasi bersih diproses menggunakan *Feature Engineering* dengan mengonversi fitur biner *SeniorCitizen* menjadi tipe data kategorikal (*string*). Langkah ini krusial untuk menyesuaikan dengan mekanisme kerja algoritma SMOTE-NC guna menghindari bias numerik saat kloning sintesis. Hasil transformasi tipe data fitur disajikan pada Tabel 3.

Tabel 3. Tipe Data Fitur Sebelum dan Sesudah Feature Engineering

No	Nama Atribut	Tipe Data Awal	Tipe Data Akhir
1	customerID	Object	DIHAPUS (Drop)
2	gender	Object	object
3	SeniorCitizen	Int64	object
4	Partner	Object	object
5	Dependents	Object	object
6	tenure	Int64	int64
7	PhoneService	Object	object
8	MultipleLines	Object	object
9	InternetService	Object	object
10	OnlineSecurity	Object	object
11	OnlineBackup	Object	object
12	DeviceProtection	Object	object
13	TechSupport	Object	object
14	StreamingTV	Object	object
15	StreamingMovies	Object	object
16	Contract	Object	object
17	PaperlessBilling	Object	object
18	PaymentMethod	Object	object
19	MonthlyCharges	Float64	float64
20	TotalCharges	Object*	float64
21	Churn	Object	object

Data yang telah ditransformasi pada Tabel 3 kemudian dibagi menjadi Data Latih 80% dan Data Uji 20% secara berstrata. Pemodelan XGBoost dieksekusi dalam dua skenario paralel, yaitu Skenario 1 menggunakan data latih asli dan Skenario 2 menggunakan data latih hasil penyeimbangan SMOTE-NC. Kedua skenario dioptimasi secara independen menggunakan hyperparameter tuning berbasis Optuna 50 iterasi. Proporsi sebaran kelas target pada data latih untuk kedua skenario divisualisasikan secara berdampingan pada Gambar 2.



Gambar 2. Distribusi Kelas Data Latih Skenario 1 dan Skenario 2 SMOTE-NC

Gambar 2 menunjukkan bahwa intervensi SMOTE-NC pada Skenario 2 berhasil mengubah struktur data yang awalnya sangat timpang menjadi seimbang (50:50). Data latih yang telah seimbang inilah yang menjadi dasar bagi model XGBoost Skenario 2 untuk belajar mengenali pola kelas minoritas.

3.2 Optimasi Hyperparameter Menggunakan Optuna

Untuk memastikan komparasi performa antarskenario berlangsung secara adil (*apple-to-apple*), kedua model XGBoost tidak dilatih menggunakan parameter bawaan (*default*), melainkan dioptimasi secara independen melalui kerangka kerja Optuna sebanyak 50 iterasi pencarian. Langkah ini memastikan masing-masing model telah mencapai potensi prediksi maksimalnya sebelum dievaluasi lebih lanjut. Hasil kombinasi parameter terbaik yang disintesis oleh Optuna untuk model dengan Data Latih Asli dan model dengan SMOTE-NC dirangkum pada Tabel 4.

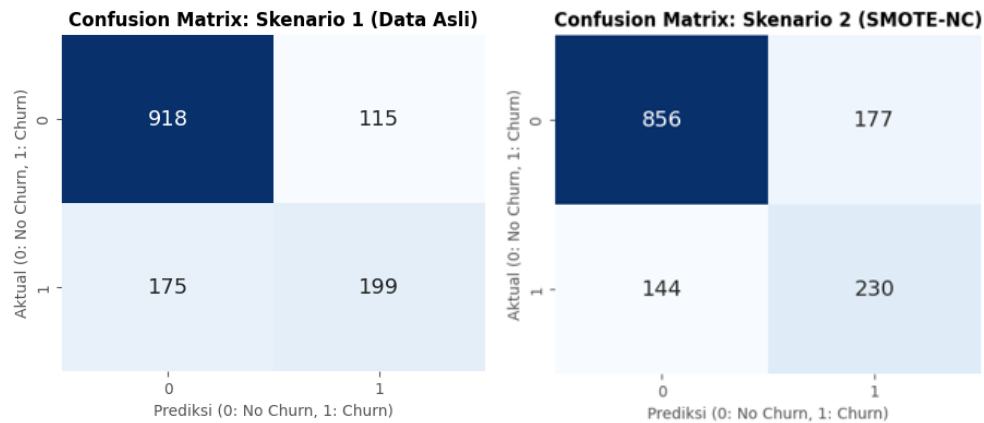
Tabel 4. Hasil Hyperparameter Tuning Optuna Kedua Skenario

Hyperparameter	Model Skenario 1 (Data Latih Asli)	Model Skenario 2 (SMOTE-NC)
n_estimators	936	250
learning_rate	0.16	0.045
max_depth	3	10
subsample	0.95	0.70
gamma	2.89	0.27
colsample_bytree	0.68	0.80

Table 4 menunjukkan bahwa Optuna mengadaptasi strategi pembelajaran yang kontras untuk kedua skenario. Pada Skenario 1 (Data Asli), model bertindak konservatif guna mencegah overfitting pada kelas mayoritas dengan membatasi kedalaman pohon ($\text{max_depth} = 3$) dan menerapkan penalti pemotongan cabang yang tinggi ($\text{gamma} = 2.89$), yang dikompensasi oleh jumlah iterasi yang tinggi ($\text{n_estimators} = 936$). Sebaliknya, pada Skenario 2 (SMOTE-NC), kompleksitas batas keputusan akibat injeksi data sintesis direspons dengan struktur pohon yang jauh lebih dalam dan bebas ($\text{max_depth} = 10$; $\text{gamma} = 0.27$). Untuk meredam noise dari agresivitas tersebut, Optuna secara logis memperlambat laju pembelajaran ($\text{learning_rate} = 0.045$), menurunkan rasio subsample menjadi 0.70, dan memangkas jumlah iterasi menjadi 250.

3.3 Evaluasi Kinerja Model XGBoost

Kinerja klasifikasi dari kedua model diuji menggunakan himpunan data uji murni sebanyak 1.407 observasi. Tingkat akurasi dan distribusi kesalahan prediksi direkam secara visual melalui *Confusion Matrix* pada Gambar 3 menggunakan *Seed 42* sebagai sampel visualisasi representatif.



Gambar 3. Confusion Matrix Hasil Pengujian Skenario 1 dan Skenario 2 (Representasi Visual pada Seed 42)

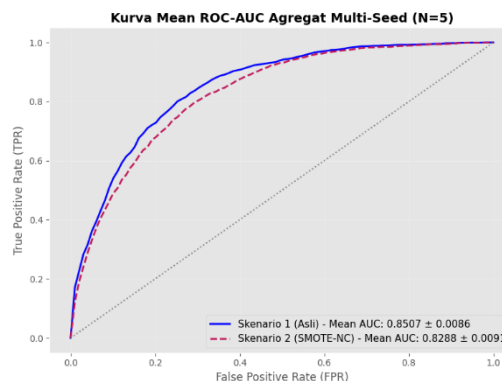
Gambar 3 menunjukkan adanya pergeseran distribusi prediksi. Pada Skenario 2, model berhasil menangkap lebih banyak kelas *churn* yang sebenarnya (*True Positive* meningkat), diiringi dengan penurunan kegagalan deteksi pelanggan *churn* (*False Negative* berkurang).

Guna mengisolasi bias variansi acak dari pembagian data uji tunggal serta menjamin validitas statistik, seluruh skenario dieksekusi secara berulang pada 5 *random seeds* berbeda. Ringkasan komparasi metrik evaluasi dalam bentuk rata-rata dan simpangan baku disajikan pada Tabel 5.

Tabel 5. Perbandingan Metrik Kinerja Klasifikasi XGBoost Multi-Seed

Metrik Evaluasi	Skenario 1 (Data Asli)	Skenario 2 (SMOTE-NC)	Skenario Kontrol (<i>Ablation Study</i>)	Dampak Intervensi (S2 vs S1)
Accuracy	0.8033 ± 0.0057	0.7795 ± 0.0080	0.7757 ± 0.0051	-2.38%
Recall (Churn)	0.5225 ± 0.0097	0.6080 ± 0.0079	0.7187 ± 0.0119	+8.55%
F1-Score (Churn)	0.5854 ± 0.0053	0.5946 ± 0.0101	0.6301 ± 0.0046	+0.92%
ROC-AUC	0.8507 ± 0.0086	0.8288 ± 0.0091	0.8423 ± 0.0065	-2.19%

Tabel 5 memperlihatkan bahwa intervensi SMOTE-NC secara konsisten meningkatkan kepekaan model dalam mendeteksi kelas minoritas (*Recall*) sebesar +8.55% dari 0.5225 ± 0.0097 menjadi 0.6080 ± 0.0079 pada seluruh pembagian data uji. Rendahnya nilai simpangan baku ($\sigma \leq 0.01$) mengonfirmasi bahwa peningkatan kepekaan deteksi ini bersifat stabil dan valid secara statistik. Evaluasi pemisahan probabilitas secara visual ditunjukkan melalui kurva ROC-AUC pada Gambar 4.



Gambar 4. Kurva Mean ROC-AUC Agregat Skenario 1 dan Skenario 2

Gambar 4 menunjukkan perbandingan kurva *Mean ROC-AUC* yang dihitung secara agregat lintas 5 *random seeds* ($N = 5$). Terlihat adanya sedikit penurunan nilai luasan kurva ROC-AUC dari

0.8507 ± 0.0086 (Skenario 1) menjadi 0.8288 ± 0.0091 (Skenario 2), atau berkurang sebesar -0.0219 (-2.19%) pasca-penyeimbangan data. Penurunan metrik global akurasi dan ROC-AUC ini mengindikasikan bahwa data sintesis SMOTE-NC memasukkan kepadatan sampel tambahan pada daerah batas keputusan, yang berdampak pada meningkatnya kesalahan identifikasi kelas mayoritas (*False Positive*).

3.4 Analisis Stabilitas Fitur Penting SHAP

Dalam mengevaluasi tingkat stabilitas peringkat fitur menggunakan koefisien *Spearman Rank Correlation*, interpretasi terhadap kekuatan hubungan didasarkan pada kriteria standar statistik yang disajikan pada Tabel 6. Standar acuan ini digunakan sebagai indikator objektif untuk menilai seberapa besar tingkat pergeseran transparansi model setelah data sintesis diinjeksikan.

Tabel 6. Kriteria Interpretasi Koefisien Korelasi Spearman

Interval Koefisien	Tingkat Kekuatan Hubungan
0.00 – 0.19	Sangat Lemah
0.20 – 0.39	Lemah
0.40 – 0.59	Sedang
0.60 – 0.79	Kuat
0.80 – 1.00	Sangat Kuat

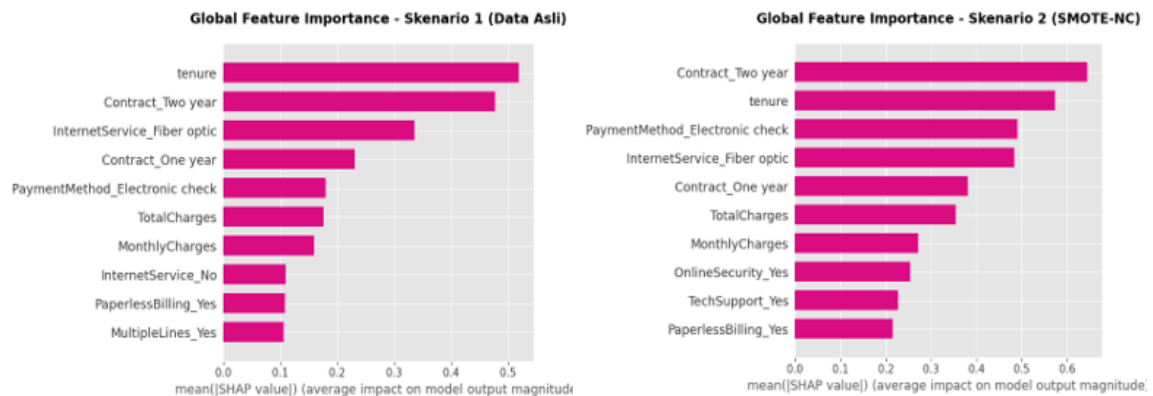
Guna memastikan bahwa analisis stabilitas tidak terikat pada satu batasan ukuran fitur, uji ketahanan (*robustness check*) dilakukan dengan mengamati variasi ambang batas parameter k ($k = 5, 10, 15$). Hasil perhitungan metrik untuk ketiga variasi tersebut disajikan pada Tabel 7.

Tabel 7. Matriks Evaluasi Stabilitas SHAP pada Variasi k Seed 42

Ambang Batas	Jaccard Similarity	Spearman Correlation	Fitur Bertahan
$k = 5$	1.0000	0.6000	5
$k = 10$	0.6667	0.8526	8
$k = 15$	1.0000	0.8000	15

Tabel 7 menunjukkan pengujian pada $k = 5$ dan $k = 15$ menghasilkan nilai *Jaccard Similarity* sebesar 1.0000, yang menunjukkan bahwa konstituen fitur sepenuhnya identik antara model *baseline* dan model SMOTE-NC pada ambang batas tersebut. Namun, koefisien Spearman pada $k = 5$ (0.6000) dan $k = 15$ (0.8000) mengindikasikan adanya rekonstruksi urutan peringkat di dalam kelompok fitur itu sendiri. Sementara itu, pada rentang interpretasi operasional ($k = 10$), nilai *Jaccard Similarity* mengalami penurunan menjadi 0.6667. Hal ini menandakan adanya 2 fitur baru yang berhasil mendominasi kelompok 10 besar dan menggeser 2 fitur terdahulu, meskipun koefisien Spearman tetap menunjukkan korelasi struktural yang sangat kuat sebesar 0.8526. Konsistensi pergeseran metrik lintas parameter ini membuktikan bahwa pengaruh intervensi SMOTE-NC terhadap logika internal model XGBoost bersifat komprehensif, bukan fluktuasi acak akibat penentuan ukuran k .

Untuk memahami mekanisme perubahan logika prediktif tersebut secara lebih mendalam, komparasi pergeseran 10 fitur teratas (*Top-10*) divisualisasikan melalui grafik *Global Feature Importance* pada Gambar 5 berikut.



Gambar 5. Fitur Top 10 SHAP Skenario 1 Data Asli dan Skenario 2 SMOTE-NC (Representasi Visual pada Seed 42)

Gambar 5 menunjukkan perbandingan pada kedua skenario, dimana pergeseran fitur tidak terjadi secara kebetulan, melainkan karena adanya perubahan distribusi probabilitas bersama (*joint distribution*) pada ruang data akibat injeksi data sintetis SMOTE-NC, yang berdampak langsung pada mekanisme percabangan (*splitting*) XGBoost.

Pada peringkat teratas, fitur numerik *tenure* dan fitur kategorikal *Contract_Two year* mengalami pertukaran posisi. Ketika SMOTE-NC memperpadat area minoritas melalui interpolasi, distribusi sampel bergeser sehingga perolehan informasi (*information gain*) dari pemisahan berdasarkan profil kontrak (*Contract_Two year*) menjadi lebih diskriminatif pada *node* awal pohon keputusan dibandingkan dengan masa berlangganan pelanggan (*tenure*).

Perubahan paling jelas terlihat pada masuknya fitur kategorikal *OnlineSecurity_Yes* dan *TechSupport_Yes* ke peringkat 8 dan 9, mengeliminasi *InternetService_No* dan *MultipleLines_Yes* dari sepuluh besar. Secara mekanis, SMOTE-NC menetapkan nilai fitur kategorikal berdasarkan modus tetangga terdekat (*k-Nearest Neighbors*). Hal ini meningkatkan homogenitas pola tertentu pada kelas minoritas buatan. XGBoost kemudian mendeteksi ketiadaan layanan dukungan teknis dan keamanan pada kelas *churn* hasil sintesis tersebut sebagai pola yang sangat dominan, sehingga frekuensi percabangan (*splits*) pada fitur-fitur tersebut meningkat secara signifikan dan mendongkrak rata-rata nilai SHAP.

Sebaliknya, fitur kontinu seperti *TotalCharges* dan *MonthlyCharges* menunjukkan stabilitas posisi secara absolut di peringkat 6 dan 7. Stabilitas ini mengindikasikan bahwa interpolasi linear pada fitur numerik mampu mempertahankan struktur varians aslinya. Analisis ini secara keseluruhan menyimpulkan bahwa manipulasi data campuran memicu distorsi interpretasi yang jauh lebih sensitif pada representasi fitur kategorikal dibandingkan fitur kontinu dalam arsitektur berbasis *tree-ensemble*.

Untuk memastikan bahwa pergeseran atributif ini murni dipicu oleh manipulasi data SMOTE-NC dan bukan akibat bias konfigurasi *hyperparameter* yang berbeda, uji validitas *multi-seed* ($N = 5$) dilakukan terhadap Skenario Kontrol (*Ablation Study*). Perbandingan hasil pengujian statistik *Jaccard Index* ($k = 10$) disajikan pada Tabel 8.

Tabel 8. Perbandingan Stabilitas Eksplanasi SHAP Multi-Seed

Skenario Pengujian	Intervensi Data	Konfigurasi Hyperparameter	Jaccard Index
Skenario 2 (SMOTE-NC)	SMOTE-NC	Hasil Optimasi Optuna	0.7939 ± 0.1233
Skenario Kontrol (Ablation Study)	SMOTE-NC	Baseline (Identik Skenario 1)	0.6713 ± 0.0886

Tabel 8 menunjukkan bahwa stabilitas *Jaccard Index* pada Skenario 2 tetap bertahan tinggi di angka 0.7939 ± 0.1233 . Sebaliknya, pada Skenario Kontrol (di mana *hyperparameter* dikunci identik dengan *baseline* Skenario 1), nilai *Jaccard Index* mengalami penurunan signifikan menjadi 0.6713 ± 0.0886 . Bukti empiris ini mengonfirmasi dua hal penting, pergeseran relevansi fitur SHAP terbukti

secara mendasar dipicu oleh perubahan ruang data akibat SMOTE-NC, dan optimasi *hyperparameter* melalui Optuna terbukti berfungsi efektif dalam memulihkan serta menjaga stabilitas interpretasi fitur model.

3.5 Analisis *Trade-off*

Berdasarkan hasil pengujian yang telah dilakukan, penanganan ketidakseimbangan kelas menggunakan metode SMOTE-NC tidak dapat dipandang sebagai solusi mutlak tanpa kompromi. Eksperimen ini mengungkap adanya fenomena *trade-off* yang terjadi pada dua aspek utama yaitu metrik performa prediktif dan keandalan interpretasi model (*Explainable AI*).

Secara ringkas, kompromi dan *trade-off* antarskenario pemodelan dirangkum dalam beberapa poin berikut:

- Peningkatan Kepekaan Klasifikasi vs Kompromi Metrik Global. Injeksi data sintesis melalui SMOTE-NC secara signifikan menyelesaikan masalah kegagalan prediksi kelas minoritas pada model baseline. Hal ini dibuktikan oleh peningkatan rata-rata Recall sebesar +8.55%. Namun, peningkatan kepekaan deteksi churn ini harus dibayar dengan kompromi berupa sedikit penurunan metrik global, seperti akurasi (-2.38%) dan ROC-AUC (-2.19%), akibat meluasnya batas keputusan (decision boundary) model dalam mentoleransi sampel sintesis.
- Optimalisasi Prediksi vs Distorsi Transparansi Fitur. Hasil analisis stabilitas menunjukkan bahwa interpretasi model mengalami pergeseran logis yang bersifat struktural. Pada ambang batas operasional $k = 10$, nilai rerata Jaccard Similarity sebesar 0.7939 ± 0.1233 mengonfirmasi bahwa mayoritas fitur utama sekitar 80% tetap bertahan secara substansial. Kendati demikian, masuknya fitur baru seperti *OnlineSecurity_Yes* dan *TechSupport_Yes* serta pertukaran posisi antara *tenure* dan *Contract_Two year* menjadi bukti konkret adanya distorsi interpretasi yang dipicu oleh homogenisasi fitur kategorikal oleh algoritma SMOTE-NC.
- Peran Optimasi Hyperparameter dalam Memitigasi Degradasi Stabilitas. Perbandingan terhadap Skenario Kontrol (Ablation Study) membuktikan bahwa injeksi SMOTE-NC tanpa penyesuaian hyperparameter berisiko menurunkan stabilitas interpretasi fitur hingga ke tingkat 0.6713 ± 0.0886 . Optimasi hyperparameter melalui Optuna pada Skenario 2 terbukti berfungsi sebagai elemen penyeimbang (*trade-off mitigator*) yang memulihkan stabilitas eksplanasi SHAP kembali ke angka 0.7939 ± 0.1233 .
- Ketahanan Lintas Ambang Batas (k). Uji ketahanan lintas parameter $k \in \{5, 10, 15\}$ membuktikan bahwa model tetap mampu menjaga konsistensi alur logika. Hal ini divalidasi oleh nilai Spearman Rank Correlation yang tetap berada pada rentang kuat hingga sangat kuat (0.6000 hingga 0.8526). Artinya, secara keseluruhan, model XGBoost yang dilatih dengan data SMOTE-NC tidak kehilangan arah prediktifnya secara radikal.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Penelitian ini secara empiris mengevaluasi dampak penanganan ketidakseimbangan kelas menggunakan SMOTE-NC pada model XGBoost yang dioptimasi dengan Optuna, sekaligus mengamati stabilitas eksplanasi fitur berbasis Explainable AI (XAI). Melalui pengujian multi-seed ($N = 5$), penerapan SMOTE-NC terbukti efektif meningkatkan kepekaan deteksi kelas minoritas (Recall) sebesar +8.55% (dari 0.5225 ± 0.0097 menjadi 0.6080 ± 0.0079) dengan kompromi penurunan akurasi global sebesar -2.38% (dari 0.8033 ± 0.0057 menjadi 0.7795 ± 0.0080). Temuan ini memberikan bukti empiris bahwa SMOTE-NC mampu mempertahankan pola logika prediktif XGBoost pada data telekomunikasi secara konsisten lintas pengacakan sampel, yang divalidasi oleh rerata nilai Jaccard Index sebesar 0.7939 ± 0.1233 ($k = 10$) serta korelasi hierarki yang sangat kuat (Spearman Correlation 0.8526). Selain merumuskan kerangka pengukuran stabilitas XAI berbasis TreeSHAP yang objektif dan dapat direplikasi untuk mengevaluasi fenomena *trade-off* penyeimbangan data campuran, penelitian ini menyusun rekomendasi intervensi bisnis yang berfokus pada profil jenis kontrak (*Contract_Two year*, *Contract_One year*), masa berlangganan (*tenure*), dan metode pembayaran (*PaymentMethod_Electronic check*) sebagai prediktor utama penentu probabilitas churn.

4.2 Saran

Berdasarkan temuan pada penelitian ini, arah pengembangan untuk penelitian mendatang yang direkomendasikan adalah sebagai berikut:

1. Pengujian pada data dinamis, menguji kerangka pengukuran stabilitas XAI ini menggunakan aliran data waktu nyata (*real-time data streaming*) untuk mengevaluasi sejauh mana pergeseran nilai SHAP berevolusi mengikuti perubahan tren perilaku pelanggan yang dinamis.
2. Komparasi metode XAI dilakukan dengan mengintegrasikan metode eksplanasi model-agnostik lainnya, seperti LIME (*Local Interpretable Model-agnostic Explanations*), untuk memastikan apakah konsistensi pergeseran logika prediktif tetap serupa di berbagai metode XAI.
3. Eksplorasi algoritma *ensemble* hibrida, menggabungkan XGBoost dengan teknik *ensemble* yang sudah terintegrasi langsung dengan mekanisme penyeimbangan data, seperti SMOTE-Boost atau RUSBoost, guna menganalisis batas toleransi *trade-off* antara performa klasifikasi dan transparansi fitur secara lebih mendalam.

5. DAFTAR PUSTAKA

- [1] H. Ribeiro, B. Barbosa, A. C. Moreira, and R. G. Rodrigues, "Determinants of churn in telecommunication services: a systematic literature review," *Management Review Quarterly*, vol. 74, no. 3, pp. 1327–1364, Sep. 2024, doi: 10.1007/s11301-023-00335-7.
- [2] A. Dhini and M. Fauzan, "Predicting Customer Churn using ensemble learning: Case Study of a Fixed Broadband Company," *International Journal of Technology*, vol. 12, no. 5, pp. 1030–1037, 2021, doi: 10.14716/ijtech.v12i5.5223.
- [3] Ibrahim Adedeji Adeniran, Christianah Pelumi Efunniyi, Olajide Soji Osundare, and Angela Omozele Abhulimen, "Implementing machine learning techniques for customer retention and churn prediction in telecommunications," *Computer Science & IT Research Journal*, vol. 5, no. 8, pp. 2011–2025, Aug. 2024, doi: 10.51594/csitrj.v5i8.1489.
- [4] R. Suguna, J. Suriya Prakash, H. Aditya Pai, T. R. Mahesh, V. Vinoth Kumar, and T. E. Yimer, "Mitigating class imbalance in churn prediction with ensemble methods and SMOTE," *Sci. Rep.*, vol. 15, no. 1, p. 16256, May 2025, doi: 10.1038/s41598-025-01031-0.
- [5] R. Rofik and J. Unjung, "Evaluation of Ridge Classifier and Logistic Regression for Customer Churn Prediction on Imbalanced Telecommunication Data," *Scientific Journal of Informatics*, vol. 12, no. 2, pp. 311–326, Jul. 2025, doi: 10.15294/sji.v12i2.24620.
- [6] J. Fonseca and F. Bacao, "Geometric SMOTE for imbalanced datasets with nominal and continuous features," *Expert Syst. Appl.*, vol. 234, p. 121053, Dec. 2023, doi: 10.1016/j.eswa.2023.121053.
- [7] N. Ahmad, M. J. Awan, H. Nobanee, A. M. Zain, A. Naseem, and A. Mahmoud, "Customer Personality Analysis for Churn Prediction Using Hybrid Ensemble Models and Class Balancing Techniques," *IEEE Access*, vol. 12, pp. 1865–1879, 2024, doi: 10.1109/ACCESS.2023.3334641.
- [8] S. Ouf, K. T. Mahmoud, and M. A. Abdel-Fattah, "A proposed hybrid framework to improve the accuracy of customer churn prediction in telecom industry," *J. Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00922-9.
- [9] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, "Customer churn prediction in telecom sector using machine learning techniques," *Results in Control and Optimization*, vol. 14, Mar. 2024, doi: 10.1016/j.rico.2023.100342.
- [10] R. Krishna, D. Jayanthi, D. S. Shylu Sam, K. Kavitha, N. K. Maurya, and T. Benil, "Application of machine learning techniques for churn prediction in the telecom business," *Results in Engineering*, vol. 24, Dec. 2024, doi: 10.1016/j.rineng.2024.103165.
- [11] S. M. Shrestha and A. Shakya, "A Customer Churn Prediction Model using XGBoost for the Telecommunication Industry in Nepal," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 652–661. doi: 10.1016/j.procs.2022.12.067.
- [12] V. Hassija *et al.*, "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," *Cognit. Comput.*, vol. 16, no. 1, pp. 45–74, Jan. 2024, doi: 10.1007/s12559-023-10179-8.

- [13] M. Imani and H. R. Arabnia, "Hyperparameter Optimization and Combined Data Sampling Techniques in Machine Learning for Customer Churn Prediction: A Comparative Analysis," *Technologies (Basel)*, vol. 11, no. 6, p. 167, Nov. 2023, doi: 10.3390/technologies11060167.
- [14] Andy Hermawan, Nila Rusiardi Jayanti, Zia Tabaruk, Faizal Lutfi Yoga Triadi, Aji Saputra, and M. Rahmat Hidayat Syachrudin, "Membangun Model Prediksi Churn Pelanggan yang Akurat (Studi Kasus tentang TELCO Company)," *Merkurius : Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 6, pp. 67–81, Oct. 2024, doi: 10.61132/mercurius.v2i6.398.
- [15] D. Asif, M. S. Arif, and A. Mukheimer, "A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry," *Results in Engineering*, vol. 26, Jun. 2025, doi: 10.1016/j.rineng.2025.104629.
- [16] A. El Attar and M. El-Hajj, "Explainable AI-driven customer churn prediction: a multi-model ensemble approach with SHAP-based feature analysis," *Front. Artif. Intell.*, vol. 9, Feb. 2026, doi: 10.3389/frai.2026.1748799.
- [17] M. Alkhozai, "An Explainable Credit Card Fraud Detection Model using Machine Learning and Deep Learning Approaches," *Journal of Applied Data Sciences*, vol. 6, no. 4, pp. 2838–2857, Dec. 2025, doi: 10.47738/jads.v6i4.962.
- [18] S. Mammadov, I. Abdulrashid, and J.-B. Sheu, "Managing Trust in AI Systems: Evidence From Controlled Evaluation of Explanation Stability Under Data Imbalance," *IEEE Trans. Eng. Manag.*, vol. 73, pp. 3083–3096, 2026, doi: 10.1109/TEM.2026.3687235.
- [19] M. KIM, "Does Class Imbalance Undermine SHAP Explanations? A Multi-Domain Empirical Study of Feature Attribution Stability," *Research Square [Preprint]*, Apr. 2026, doi: 10.21203/rs.3.rs-9489204/v1.
- [20] BlastChar, "Dataset Teleco Customer Churn IBM," Kaggle.com. Accessed: Jun. 30, 2026. [Online]. Available: <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>
- [21] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," *Preprint*, Aug. 2019.
- [22] V. Sinap, "Bankruptcy Prediction with Optuna-Enhanced Ensemble Machine Learning Methods: A Comparison of Oversampling and Undersampling Techniques," *DÜMF Mühendislik Dergisi*, vol. 16, no. 1, pp. 97–113, Mar. 2025, doi: 10.24012/dumf.1597564.
- [23] S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," *IEEE Access*, vol. 9, pp. 62118–62136, 2021, doi: 10.1109/ACCESS.2021.3073776.
- [24] V. Verma and R. K. Aggarwal, "A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, p. 43, Dec. 2020, doi: 10.1007/s13278-020-00660-9.
- [25] P. Schober, C. Boer, and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation," *Anesth. Analg.*, vol. 126, no. 5, pp. 1763–1768, May 2018, doi: 10.1213/ANE.0000000000002864.
- [26] W. Pirie, "Spearman Rank Correlation Coefficient," in *Encyclopedia of Statistical Sciences*, Wiley, 2005. doi: 10.1002/0471667196.ess2499.pub2.