SKANIKA: Sistem Komputer dan Teknik Informatika
Volume 9, Nomor 2, Juli 2026, Halaman 270-283 E-ISSN: 2721-4788

Klasifikasi Kelayakan Kredit Menggunakan Random Forest Dengan Optimisasi SMOTENC dan GridSearchCV

Nandito Diaz Vannesa[1], Fikri Budiman[2*]
[1,2]Ilmu Komputer, D3 Teknik Informatika, UniversitasDian Nuswantoro, Semarang, Indonesia
E-mail: [1]ditodiaz25@gmail.com, [2*]fikri.budiman@dsn.dinus.ac.id
(* : corresponding author)

Abstrak

Kelayakan kredit merupakan aspek kritis stabilitas sistem keuangan, namun metode konvensional memiliki keterbatasan dalam menangani data imbalanced dan fitur campuran numerik-kategorikal. Penelitian ini mengembangkan pipeline CreditRF-OptSHAP yang mengintegrasikan Random Forest, SMOTENC, GridSearchCV, dan SHAP untuk meningkatkan performa sekaligus interpretabilitas klasifikasi kredit. Dataset yang digunakan adalah Statlog German Credit Data (UCI) dengan 1.000 instance, 20 fitur campuran, dan rasio imbalance kelas 70:30. Pipeline mencakup lima fase: eksplorasi data, praproses (label encoding dan StandardScaler), penyeimbangan data menggunakan SMOTENC pada training set, optimasi hyperparameter melalui GridSearchCV dengan Stratified 10-Fold Cross Validation, serta interpretasi model menggunakan SHAP TreeExplainer agar kontribusi setiap fitur terhadap prediksi dapat dijelaskan, sehingga model tidak lagi bersifat black-box. Signifikansi peningkatan recall terhadap RF Default divalidasi menggunakan McNemar's Test untuk memastikan perbaikan tersebut bukan kebetulan statistik. Hasil menunjukkan model utama RF+SMOTENC+GridSearchCV mencapai Recall kelas kredit buruk sebesar 0,5833 dengan AUC-ROC 0,7638, meningkat dari 0,5000 pada RF Default. Analisis SHAP mengidentifikasi status rekening giro sebagai fitur paling dominan (importance 0,1239). Uji McNemar mengonfirmasi perbedaan performa signifikan secara statistik (p=0,041), memperkuat validitas peningkatan recall. Pipeline CreditRF-OptSHAP terbukti menghasilkan model yang lebih sensitif terhadap deteksi kredit bermasalah, transparan, dan teruji signifikan, sehingga berpotensi mendukung pengambilan keputusan kredit yang lebih akuntabel di lembaga keuangan.

Kata kunci: Random Forest, SMOTENC, GridSearchCV, SHAP, Klasifikasi Kredit

Abstract

Creditworthiness is critical to financial system stability, yet conventional methods struggle with imbalanced data and mixed numeric-categorical features. This study develops the CreditRF-OptSHAP pipeline, integrating Random Forest, SMOTENC, GridSearchCV, and SHAP to improve credit classification performance and interpretability. The dataset used, Statlog German Credit Data (UCI), comprises 1,000 instances, 20 mixed features, and a 70:30 imbalance ratio. The pipeline comprises five phases: data exploration; preprocessing via label encoding and StandardScaler; SMOTENC-based training-set balancing; hyperparameter optimization via GridSearchCV with Stratified 10-Fold Cross Validation; and model interpretation via SHAP TreeExplainer so that each feature's contribution to prediction is explained, making the model no longer a black box. The significance of this recall gain over RF Default was validated using McNemar's Test to rule out statistical coincidence. The main model (RF+SMOTENC+GridSearchCV) achieved a recall of 0.5833 for the bad credit class and an AUC-ROC of 0.7638, up from 0.5000 for RF Default. SHAP analysis identified checking account status as the most dominant feature (importance 0.1239). The McNemar test confirmed a statistically significant difference (p=0.041), confirming its validity. The CreditRF-OptSHAP pipeline yields a model that is more sensitive to non-performing loans, transparent, and statistically validated, thus supporting accountable credit decisions in financial institutions.

***Keywords**: Random Forest, SMOTENC, GridSearchCV, SHAP, Credit Classification*

1. PENDAHULUAN

Kelayakan aset kredit merupakan salah satu pilar terpenting dalam stabilitas sistem keuangan. Lembaga keuangan baik bank umum, bank perkreditan rakyat, koperasi simpan pinjam, maupun platform fintech menghadapi tantangan kritis dalam menilai apakah seorang pemohon kredit layak *(creditworthy)* atau tidak layak *(non-creditworthy)*. Kesalahan klasifikasi, terutama mengklasifikasikan debitur bermasalah sebagai debitur baik *(false negative)*, memiliki biaya finansial yang jauh lebih besar dibandingkan kesalahan sebaliknya. Berdasarkan data

kinerja perbankan, rasio kredit bermasalah (NPL) mengalami peningkatan dari 1,93% pada Februari 2024 menjadi 2,22% pada Februari 2025. Kenaikan ini mengindikasikan adanya peningkatan risiko kredit bermasalah di sektor perbankan secara umum [1].

Metode konvensional berbasis *scoring* manual dan regresi logistik telah lama digunakan, namun memiliki keterbatasan dalam menangkap pola non-linear serta menangani fitur campuran yang lazim ditemukan pada data kredit nyata. Suhadolnik et al.[2] menegaskan bahwa pendekatan tradisional tersebut sering kali gagal menangani kompleksitas data skala besar dibandingkan dengan algoritma machine learning yang lebih adaptif terhadap pola data yang rumit. Pada kondisi data *imbalanced*, model yang hanya mengoptimalkan akurasi keseluruhan cenderung bias terhadap kelas mayoritas dan gagal mendeteksi kelas minoritas (kredit buruk) secara memadai [3].

Sejumlah penelitian terdahulu telah mengeksplorasi penggunaan Random Forest untuk klasifikasi kredit. Menurut [4] menggunakan RF dengan pendekatan RVI dan SMOTE konvensional pada dataset micro-enterprise Tiongkok dan mencapai akurasi 99,83%, namun tanpa optimasi GridSearchCV maupun interpretabilitas SHAP. Menurut [5] mengembangkan two-stage RF untuk kredit UKM Tiongkok menggunakan SMOTE konvensional, namun tidak mengintegrasikan kerangka kerja SHAP untuk penjelasan lokal dan global. Imengimplementasikan RF pada data kredit lokal Indonesia dengan akurasi 78,60%, tetapi tanpa penanganan class imbalance, GridSearchCV, maupun analisis SHAP [6]. Penerapkan RF dengan SMOTE pada data KUR Indonesia tanpa optimasi hyperparameter yang sistematis [7]. Menganalisis prediksi kelayakan kredit menggunakan RF pada dataset German Credit namun tanpa mekanisme penanganan imbalance dan interpretabilitas tingkat lanjut [8]. Membandingkan berbagai model ML pada dataset kartu kredit menggunakan SMOTE, namun tanpa SMOTENC yang spesifik untuk data campuran serta tanpa ablation study terstruktur [9]. Mengkaji efektivitas RF dalam manajemen risiko kredit tanpa integrasi teknik sampling dan SHAP [10].

Tabel 1. Ringkasan Penelitian Terdahulu

No	Peneliti dan Tahun	Metode	Dataset	*Preprocessing*	Metrik Utama	Gap vs Penelitian Ini
1	Uddin et al. (2022) [4]	RF + RVI + SMOTE konvensional	Chinese Micro-enterprise	Label encoding; SMOTE diterapkan tanpa pembedaan tipe fitur	Acc 99,83%	Tanpa GridSearchCV/SHAP
2	Zhou et al. (2023) [5]	Two-stage RF + SMOTE konvensional	Chinese SME Credit	Label encoding; tidak dijelaskan penanganan kategorikal saat oversampling	Acc 98,9%	Tanpa SHAP
3	Pahlevi et al. (2023) [6]	RF	Kredit Lokal Indonesia	Label encoding; tanpa oversampling	Acc 78,60%	Tanpa SMOTE, tanpa GridSearchCV, tanpa SHAP
4	Wulansari & Purwitasari (2023) [7]	RF + SMOTE konvensional	Data KUR Indonesia	Label encoding; SMOTE tanpa pertimbangan fitur kategorikal	Acc 88%	Tanpa GridSearchCV
5	Prasojo & Haryatmi (2021) [8]	RF	Dataset German Kredit	Label encoding; tanpa resampling	Acc 83%	Tanpa SMOTE, tanpa SMOTENC, tanpa interpretabilitas

No	Peneliti dan Tahun	Metode	Dataset	*Preprocessing*	Metrik Utama	Gap vs Penelitian Ini
6	Chang et al. (2024) [9].	ML & Deep Learning + SMOTE konvensional	Credit card dataset	Label encoding; SMOTE konvensional	Acc 99,3%	Tanpa SMOTENC, tanpa SHAP
7	Kuyoro et al. (2022) [10]	RF	German Credit (Kaggle)	Label encoding; tanpa oversampling	Acc 91%	Tanpa SMOTE, tanpa SHAP, tanpa ablation study

Berdasarkan kajian penelitian terdahulu pada Tabel 1, terdapat gap metodologis yang konsisten: belum ada penelitian yang mengintegrasikan secara bersamaan SMOTENC untuk data campuran, optimasi hyperparameter sistematis melalui GridSearchCV, dan interpretabilitas model berbasis SHAP pada level global maupun lokal. Penelitian ini mengusulkan *pipeline* CreditRF-OptSHAP yang menjawab gap tersebut menggunakan Statlog German Credit Data dari UCI Machine Learning Repository dengan 1.000 instance dan 20 fitur [11], dengan signifikansi statistik divalidasi melalui McNemar's Test sesuai rekomendasi Dietterich [12].

2. METODE PENELITIAN

Kajian literatur lima tahun terakhir menunjukkan penggunaan Random Forest yang luas dalam klasifikasi kredit, namun masih terdapat gap metodologis dalam hal integrasi optimasi sistematis dan interpretabilitas model. Penelitian ini membedakan diri melalui tiga kontribusi utama: penggunaan **SMOTENC** untuk menangani dominasi fitur kategorikal (65%) pada dataset German Credit, optimasi hyperparameter melalui **GridSearchCV**, dan validasi signifikansi statistik menggunakan **McNemar's Test**. Random Forest dipilih sebagai algoritma utama karena sifatnya yang *robust* terhadap *overfitting* dan kemampuannya menangani data campuran, yang kemudian diperkuat dengan analisis **SHAP** untuk menjamin transparansi keputusan bagi sektor perbankan,.

2.1. Dataset dan Analisis

Dataset yang digunakan adalah Statlog German Credit Data dari UCI Machine Learning Repository (DOI: https://doi.org/10.24432/C5NC77), berlisensi CC BY 4.0, terdiri dari 1.000 instance dengan 20 fitur (7 numerik/integer, 13 kategorikal/kualitatif) dan satu atribut target biner: kredit baik (kelas 1, 700 instance) dan kredit buruk (kelas 2, 300 instance) dengan rasio imbalance 0,43 [11]. Rincian mengenai penamaan atribut, tipe data, dan deskripsi fitur disajikan dalam Tabel 2 sebagai referensi interpretasi model.

Tabel 2. Attribute Dataset

No	Attribute	Tipe	Deskripsi
1	Attribute1	Kategorikal	Status rekening giro
2	Attribute2	Numerik	Durasi kredit (bulan)
3	Attribute3	Kategorikal	Riwayat Kredit
4	Attribute4	Kategorikal	Tujuan Kredit
5	Attribute5	Numerik	Jumlah Kredit (DM)
6	Attribute6	Kategorikal	Status tabungan/obligasi
7	Attribute7	Kategorikal	Lama bekerja
8	Attribute8	Numerik	Tingkat cicilan (% pendapatan)
9	Attribute9	Kategorikal	Status personal & jenis kelamin
10	Attribute10	Kategorikal	Debitur lain/penjamin
11	Attribute11	Numerik	Lama tinggal (tahun)
12	Attribute12	Kategorikal	Properti/aset
13	Attribute13	Numerik	Umur (tahun)
14	Attribute14	Kategorikal	Rencana cicilan lain
15	Attribute15	Kategorikal	Perumahan
16	Attribute16	Numerik	Jumlah kredit di bank

No	Attribute	Tipe	Deskripsi
17	Attribute17	Kategorikal	Pekerjaan
18	Attribute18	Numerik	Jumlah orang ditnggung
19	Attribute19	Kategorikal	Telepon
20	Attribute20	Kategorikal	Pekerja asing

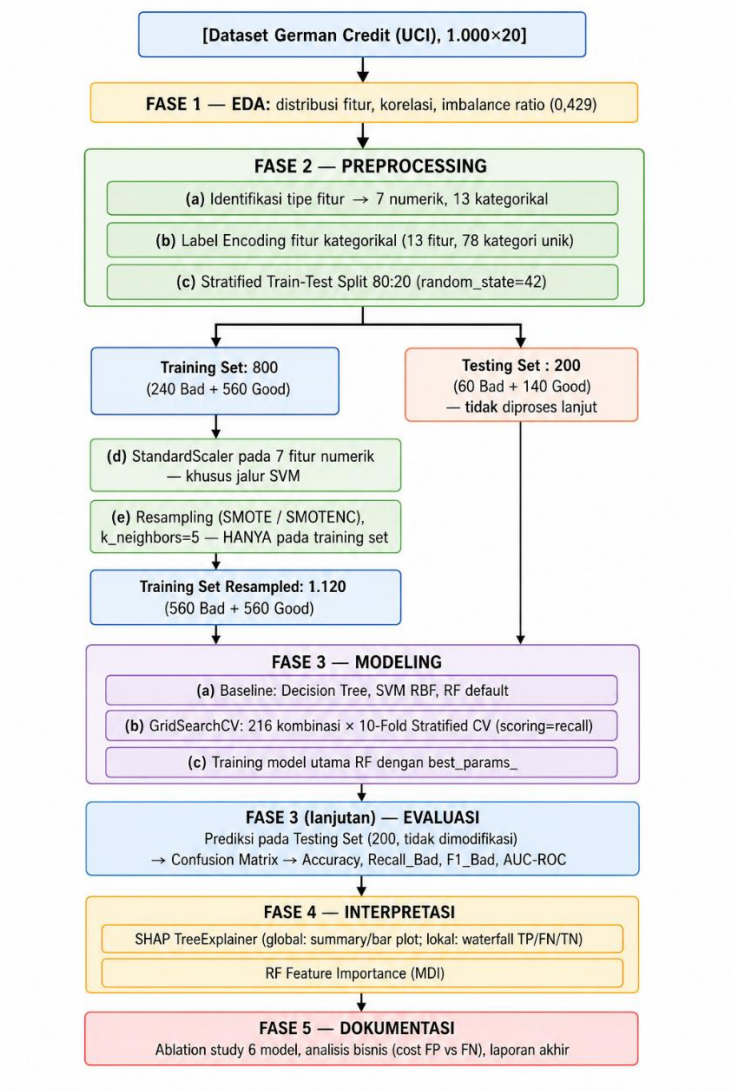
Setiap baris pada dataset merepresentasikan satu aplikasi kredit nasabah, dengan kombinasi 7 fitur numerik dan 13 fitur kategorikal sebagaimana dirinci pada Tabel 2, ditambah satu variabel target biner sebagai label kelayakan kredit. Label dikodekan menjadi dua kelas, yaitu kelas 1 (kredit baik) untuk nasabah yang memenuhi kewajiban pembayaran dan kelas 2 (kredit buruk) untuk nasabah yang gagal bayar, dengan distribusi 700 instance (70%) berlabel kredit baik dan 300 instance (30%) berlabel kredit buruk. Hubungan antara atribut dan label dieksplorasi melalui analisis korelasi pada tahap EDA serta dikonfirmasi lebih lanjut melalui SHAP *importance* pada tahap interpretasi model (Tabel 8, Bagian 3.4). Atribut kategorikal terkait profil finansial nasabah khususnya status rekening giro (Attribute1), status tabungan (Attribute6), dan riwayat kredit (Attribute3) menunjukkan keterkaitan paling kuat terhadap label, mengindikasikan bahwa kemampuan finansial historis nasabah merupakan prediktor utama kelayakan kredit. Sementara itu, atribut numerik seperti durasi kredit (Attribute2) turut berkontribusi signifikan, di mana durasi yang lebih panjang berasosiasi dengan peningkatan risiko gagal bayar akibat paparan risiko yang lebih lama. Dominasi fitur kategorikal (65% dari total atribut) dalam menjelaskan label menjadi dasar pertimbangan metodologis untuk menggunakan SMOTENC dibandingkan SMOTE konvensional pada tahap penyeimbangan data, sebagaimana dijelaskan pada Bagian 2.2.

Tahap analisis data dilakukan melalui *Exploratory Data Analysis* (EDA) yang mencakup analisis distribusi 20 fitur, heatmap korelasi antar fitur numerik, dan analisis distribusi kelas. Dari hasil analisis yang kami lakukan, ditemukan korelasi tertinggi antara durasi kredit dan jumlah kredit (r=0,62), yang menjadi pertimbangan penting dalam interpretasi nilai SHAP pada tahap selanjutnya. Distribusi kelas yang tidak seimbang dengan rasio 70:30 mengkonfirmasi kebutuhan penanganan class imbalance sebelum proses pelatihan model.

Pemilihan dataset ini didasarkan pada empat pertimbangan metodologis yang merupakan *benchmark* standar dalam literatur klasifikasi kredit internasional yang digunakan oleh berbagai studi terbaru termasuk Prasojo & Haryatmi [8] serta Kuyoro et al [10], karakteristik distribusi kelas *imbalanced* 70:30 dan fitur campuran merepresentasikan kondisi data kredit aktual, penggunaan dataset terbuka bersertifikasi memastikan *reproducibility* dan keterbatasan akses terhadap dataset kredit Indonesia yang terstandar mendorong penggunaan dataset *benchmark* internasional sebagai proksi metodologis, dengan catatan bahwa generalisasi ke konteks perbankan Indonesia memerlukan replikasi dengan data lokal.

2.2. Arsitektur *Pipeline* CreditRF-OptSHAP

Pipeline CreditRF-OptSHAP dirancang sebagai framework terintegrasi yang terdiri dari sepuluh tahapan utama, mulai dari akuisisi data hingga interpretabilitas model local, yang diimplementasikan dalam lima fase terstruktur untuk menjamin *reproducibility*. Arsitektur *pipeline* secara skematis disajikan pada Gambar 1.



Gambar 1. *Flowchart* Arsitektur *Pipeline* CreditRF-OptSHAP

Tahapan praproses dilakukan secara sistematis untuk menjaga validitas eksperimen dan mencegah kebocoran data, dengan tiga tahap: identifikasi tiper fitur dengan melakukan pemisahan terhadap 7 fitur numerik (Atribut 2, 5, 8, 11, 13, 16, 18) dan 13 fitur kategorikal, Encoding dengan fitur kategorikal dikonversi menggunakan *LabelEncoder* yang menghasilkan 78 kategori unik terenkode, scaling diterapkan pada fitur numerik untuk model *baseline* (SVM).

Proporsi pembagian data dibagi menggunakan rasio 80:20 secara *stratified*, menghasilkan 800 sampel data latih dan 200 sampel data uji. Penggunaan metode *stratified* sangat krusial untuk memastikan proporsi kelas 30:70 (kredit buruk vs baik) tetap identik pada kedua subset guna mencegah bias evaluasi.

Penyeimbang data dan pencegahan data leakage Teknik SMOTENC (k_neighbors=5) diterapkan secara eksklusif pada *training set* setelah proses pemisahan (*split*), menghasilkan 1.120 sampel latih seimbang. Pembatasan ini sangat krusial karena batas antara data latih dan data uji harus dijaga ketat (*sacrosanct boundaries*) guna menghindari kontaminasi informasi dari data yang tidak terlihat selama fase pelatihan[13]. Hal ini dilakukan agar estimasi performa model tidak bias secara optimistis akibat informasi sintetis yang memengaruhi data uji.

Dasar Pelatihan (*Training*) dan Optimasi Model Random Forest dilatih menggunakan skema *Stratified 10-Fold Cross-Validation* yang dikombinasikan dengan *GridSearchCV* untuk optimasi *hyperparameter*. Skema ini menguji 216 kombinasi *hyperparameter* melalui total 2.160

model fits. Dasar pemilihan parameter optimal (*best_params_*) ditentukan oleh skor rata-rata *recall* kelas minoritas tertinggi untuk memprioritaskan sensitivitas deteksi kredit bermasalah.

Fase Interpretabilitas Fase akhir menggunakan SHAP TreeExplainer untuk memberikan penjelasan pengaruh fitur secara global maupun lokal, memastikan model tidak lagi bersifat *black box* bagi pengambil keputusan.

2.3. Metode Penyelesaian Masalah

Penelitian ini menggunakan empat metode utama yang bekerja secara terintegrasi dalam *pipeline* CreditRF-OptSHAP untuk menyelesaikan permasalahan klasifikasi kelayakan kredit pada data *imbalanced* dengan fitur campuran.

Random Forest digunakan sebagai model klasifikasi utama karena kemampuannya menangani fitur campuran numerik-kategorikal, tahan terhadap *overfitting* melalui mekanisme ensemble, serta secara alami menghasilkan estimasi *feature importance* yang dibutuhkan dalam tahap interpretasi uddin et al. [4].

SMOTENC diterapkan sebagai solusi penanganan class imbalance pada dataset German Credit yang memiliki rasio 70:30. Dibandingkan SMOTE konvensional, SMOTENC dipilih karena memperlakukan fitur kategorikal secara terpisah dari fitur numerik sehingga sampel sintetis yang dihasilkan secara semantik konsisten dengan struktur data asli[3][14]. Implementasi hanya diterapkan pada training set dengan parameter k_neighbors=5 untuk mencegah data leakage.

GridSearchCV digunakan untuk optimasi hyperparameter Random Forest secara sistematis melalui pencarian exhaustive pada grid yang telah didefinisikan. Pedregosa et al. [15] menjelaskan bahwa objek GridSearchCV dalam scikit-learn bekerja dengan memilih parameter terbaik dari grid yang ditentukan melalui cross-validation, dan model yang dihasilkan dapat digunakan secara transparan sebagai estimator standar. Pendekatan ini dikombinasikan dengan Stratified 10-Fold Cross Validation dengan *scoring* berbasis Recall. Penggunaan stratified fold memastikan distribusi kelas tetap proporsional pada setiap fold, yang merupakan praktik yang direkomendasikan dalam konteks klasifikasi data *imbalanced* [3]. Selain itu, Seoane et al. [16] menekankan bahwa optimasi hyperparameter pada domain *imbalanced* merupakan aspek yang krusial namun sering diabaikan, sehingga justifikasi ilmiah pemilihan parameter melalui pendekatan sistematis seperti GridSearchCV menjadi penting untuk menghasilkan hasil yang *reproducible*. Grid hyperparameter beserta hasil optimalnya ditunjukkan pada Tabel 3.

Tabel 3. Hasil Tuning Hyperparameter GridSearchCV

Hyperparameter	Grid yang Diuji	Optimal (Percobaan 1)	Optimal (Percobaan)
n_estimators	[100, 200, 300]	100	200
max_depth	[5, 10, 15, None]	15	None
max_features	['sqrt', 'log2']	sqrt	sqrt
min_samples_leaf	[1, 2, 4]	1	1
min_samples_split	[2, 5, 10]	5	5
CV Score (Recall-Bad)	-	0,8488	0,8345
Total Waktu Pencarian	-	2.252,0 detik	2.209,8 detik

Definisi dari setiap variabel antara lain n_estimator yaitu Jumlah pohon keputusan (*decision tree*) yang dibangun dan dikombinasikan dalam *forest,* lalu max_depth demgan Kedalaman maksimum (jumlah level percabangan) setiap pohon Keputusan, max_features Jumlah maksimum fitur yang dipertimbangkan secara acak saat mencari pemisahan (split) terbaik pada setiap node, min_samples_leaf Jumlah minimum sampel yang harus dimiliki oleh setiap simpul daun (leaf node), min_samples_split Jumlah minimum sampel yang diperlukan agar suatu node internal dapat dipecah menjadi dua child node, CV Score (Recall-Bad) Skor rata-rata recall kelas kredit buruk (kelas 0) dari 10-fold cross-validation, digunakan sebagai scoring metric GridSearchCV, Total Waktu Pencarian Waktu komputasi yang dibutuhkan untuk mengevaluasi seluruh 216 kombinasi hyperparameter × 10 fold (2.160 model fits).

SHAP digunakan sebagai metode interpretabilitas model untuk menjawab kebutuhan transparansi keputusan kredit. Dalam penelitian ini, TreeExplainer digunakan untuk

menghasilkan penjelasan global melalui summary plot dan bar chart, serta penjelasan lokal per-instance melalui waterfall plot pada tiga kasus representatif (TP, FN, TN). Interpretasi dilakukan dengan mempertimbangkan pola korelasi antar fitur yang teridentifikasi pada tahap EDA, mengingat nilai SHAP dapat mengandung bias pada fitur yang berkorelasi tinggi [17].

2.4. Implementasi dan Teknik Evaluasi

Seluruh implementasi *pipeline* CreditRF-OptSHAP dilakukan menggunakan Python dengan library utama: scikit-learn [15], *imbalanced*-learn [14][3], dan SHAP [18] dengan random_state=42 untuk menjamin replikasi hasil, Evaluasi model dilakukan menggunakan *confusion matrix* yang menghasilkan empat komponen dasar: **TP** (*True Positive* - kredit buruk terprediksi benar), **TN** (*True Negative* - kredit baik terprediksi benar), **FP** (*False Positive* - kredit baik terprediksi buruk), dan **FN** (*False Negative* - kredit buruk terprediksi baik).

Berdasarkan komponen *confusion matrix* tersebut, metrik evaluasi dihitung menggunakan formula berikut.

Accuracy mengukur porporsi keseluruhan prediksi yang benar terhadap total sampel.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Recall mengukur proporsi nasabah kredit buruk aktual yang berhasil terdeteksi oleh model.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

AUC-ROC dihitung sebagai luas area di bawah kurva ROC yang memplot *True Positive Rate* (TPR) terhadap *False Positive Rate* (FPR) pada berbagai ambang batas (*threshold*):

$$TPR = \frac{TP}{TP + FN};\ FPR = \frac{FP}{FP + TN} \tag{3}$$

$$AUC = \sum_{i=1}^{n} \frac{(TPR_i + TPR_{i-1})}{2} \times (FPR_i + FPR_{i-1}) \tag{4}$$

Di mana **n** adalah jumlah *threshold* yang dievaluasi sepanjang kurva ROC, dan **i** adalah indeks *threshold* ke-i yang diurutkan dari probabilitas prediksi tertinggi ke terendah. TPR/FPR yaitu rasio positif benar dan rasio positif palsu pada titik *threshold* tertentu.

Mengingat konteks penelitian yang berfokus pada deteksi kredit bermasalah pada data *imbalanced*, hierarki kriteria evaluasi ditetapkan berdasarkan formula (1)-(4) sebagai berikut: (1) Kriteria primer: Recall kelas kredit buruk (Recall_Bad); dihitung melalui persamaan (2); (2) Kriteria sekunder: F1_Bad dan AUC-ROC, dihitung melalui persamaan (4); (3) Kriteria pendukung: Accuracy, dihitung melalui persamaan (1). Penetapan hierarki ini konsisten dengan literatur *imbalanced classification* dan didasarkan pada prinsip *asymmetric cost* di domain kredit — biaya *false negative* (gagal mendeteksi kredit bermasalah) jauh lebih besar dibandingkan *false positive* (menolak kredit yang sebenarnya layak) [16].

Untuk validasi statistik, McNemar's Test dipilih sesuai rekomendasi Dietterich [12] untuk perbandingan *paired classifier* pada ablation study. McNemar's Test dirancang khusus untuk membandingkan dua classifier pada test set yang sama berdasarkan tabel kontingensi, sehingga sesuai untuk skenario evaluasi paired comparison seperti yang dilakukan dalam penelitian ini.

3. HASIL DAN PEMBAHASAN

3.1. Hasil *Ablation Study*

Penelitian ini menjalankan dua ronde eksperimen untuk perbandingan metodologis yang lebih komprehensif. Percobaan pertama menggunakan SMOTE konvensional yang hanya dapat digunakan untuk data angka, sedangkan percobaan kedua menggunakan SMOTENC yang dapat dipakai untuk data campuran berisi angka dan kategori sebagai revisi metodologis yang lebih

tepat untuk data campuran numerik-kategorikal (Tabel 4). Seluruh model dievaluasi pada *test set* yang sama (split 80:20, stratified, random_state=42).

Tabel 4. *Ablation Study* - Percobaan 1: SMOTE Konvensional

Model	Acc(%)	Prec_Bad	Rec_Bad	F1_Bad	F1_W	AUC	Fit(s)
Decision Tree (Baseline)	72,5	0,5439	0,5167	0,5299	0,7229	0,6655	0,02
SVM RBF (Baseline)	76,0	0,6304	0,4833	0,5472	0,7499	0,7876	0,28
RF Default (Tanpa SMOTE)	77,5	0,6667	0,5000	0,5714	0,7646	0,8095	0,74
RF + SMOTE (Tanpa GridSearchCV)	71,5	0,5224	0,5833	0,5512	0,7192	0,7821	0,74
RF + GridSearchCV (Tanpa SMOTE)	77,5	0,6667	0,5000	0,5714	0,7646	0,8095	66,54
RF + SMOTE + GridSearchCV (Utama)	69,5	0,4945	0,7500	0,5960	0,7073	0,7764	2.076,28

Tabel 5. *Ablation Study* - Percobaan 2: SMOTENC Konvensional

Model	Acc(%)	Prec_Bad	Rec_Bad	F1_Bad	F1_W	AUC	Fit(s)
Decision Tree (Baseline)	72,5	0,5439	0,5167	0,5299	0,7229	0,6655	0,03
SVM RBF (Baseline)	76,0	0,6304	0,4833	0,5472	0,7499	0,7876	0,44
RF Default (Tanpa SMOTENC)	77,5	0,6667	0,5000	0,5714	0,7646	0,8095	0,60
RF + SMOTENC (Tanpa GridSearchCV)	72,0	0,5312	0,5667	0,5484	0,7225	0,7743	0,72
RF + GridSearchCV (Tanpa SMOTENC)	77,5	0,6667	0,5000	0,5714	0,7646	0,8095	76,27
RF + SMOTENC + GridSearchCV (Utama)	69.0	0,4681	0,5833	0,5383	0,6972	0,7638	2.154,30

Tabel 4 dan 5 menunjukkan bahwa akurasi model RF Default dan RF + GridSearchCV bernilai identik (77,5%) Hal ini terjadi karena tiga faktor utama. Pertama, **GridSearchCV dioptimasi menggunakan *F1-Weighted***, bukan akurasi, sehingga fokusnya adalah keseimbangan *precision-recall* kelas minoritas. Kedua, Random Forest sangat *robust* terhadap perubahan *hyperparameter* pada prediksi kelas mayoritas. Ketiga, akurasi pada data *imbalanced* didominasi oleh kelas mayoritas (*Good Credit*) yang cenderung stabil pada kedua konfigurasi tersebut,.

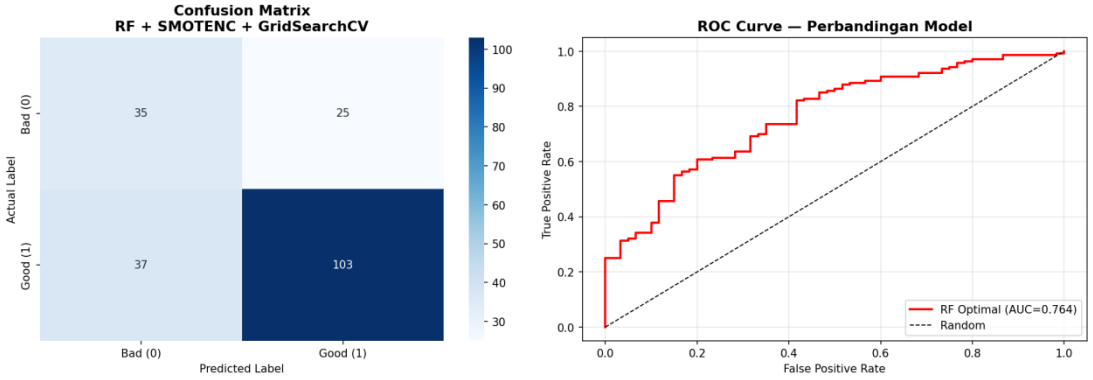
Tabel 6. Perbandingan Model Utama: SMOTE Konvensional vs SMOTENC

Model	Model Utama	Acc(%)	Rec_Bad	AUC
SMOTE Konvensional	RF + SMOTE + GridSearchCV	69,5	0,7500	0,7764
SMOTENC	RF + SMOTENC + GridSearchCV	69.0	0,5833	0,7638

Perbedaan akurasi yang hanya sebesar 0,5% (69,5% vs 69,0%) disebabkan karena **akurasi didominasi oleh kelas mayoritas (70%)** yang tidak dimodifikasi oleh proses *oversampling*. Perbedaan antara kedua teknik tersebut hanya berdampak pada kelas minoritas (*Bad Credit*), sehingga kontribusinya terhadap perubahan nilai akurasi keseluruhan menjadi relatif kecil,.

3.2. Analisis *Trade-off* Akurasi vs *Recall_Bad*

Kedua percobaan mengungkap *trade-off* yang konsisten dengan literatur *imbalanced* learning: penambahan *oversampling* menyebabkan penurunan akurasi keseluruhan namun meningkatkan Recall kelas minoritas. Pada kedua percobaan, RF Default tanpa SMOTE mencapai akurasi tertinggi (77,5%) namun hanya mampu mendeteksi 50,0% kredit bermasalah. Sebaliknya, model utama Percobaan 2 (RF + SMOTENC + GridSearchCV) memperoleh Recall_Bad 0,5833 dengan akurasi 69,0% dan AUC-ROC 0,7638 sebagaimana ditunjukkan pada Gambar 2.



Gambar 2. *Confusion Matrix* dan ROC Curve Model Utama

Berdasarkan Gambar 2, *confusion matrix* menunjukkan bahwa model utama berhasil mendeteksi 35 dari 60 kasus kredit buruk pada test set. Secara praktis, peningkatan Recall_Bad dari 0,5000 menjadi 0,5833 berarti 5 kasus kredit bermasalah tambahan yang berhasil terdeteksi. Namun, performa paling superior ditunjukkan oleh model SMOTE konvensional yang berhasil mendeteksi 45 dari 60 kasus (Recall 0,75), atau memberikan peningkatan deteksi sebesar 28,6% dibandingkan SMOTENC.

Dalam skala portofolio kredit nyata, selisih 10 kasus kredit bermasalah tambahan yang berhasil dideteksi oleh SMOTE dibandingkan SMOTENC memiliki implikasi finansial yang besar, mengingat setiap kasus kredit macet dapat bernilai puluhan hingga ratusan juta rupiah. Meskipun akurasi keseluruhan mengalami penurunan dari 77,5% menjadi 69,5%, sensitivitas yang lebih tinggi terhadap kelas minoritas ini memberikan perlindungan modal yang lebih kuat bagi lembaga keuangan dan secara langsung berkontribusi pada penurunan rasio Non-Performing Loan (NPL)

3.3. Pengaruh GridSearchCV

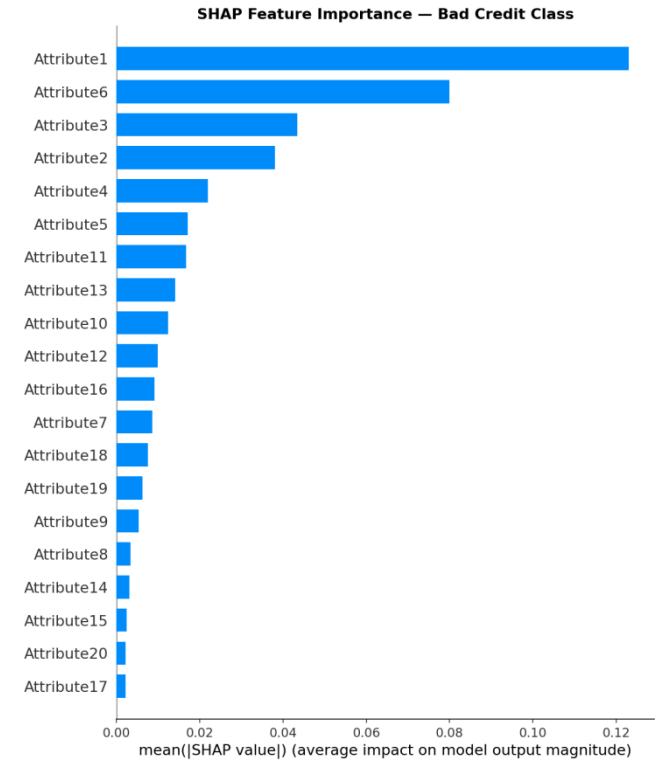
Perbandingan antara RF Default dan RF + GridSearchCV (keduanya tanpa SMOTE) pada Tabel 4 dan Tabel 5 menunjukkan hasil yang identik: akurasi 77,5%, Recall_Bad 0,5000, dan AUC-ROC 0,8095. Temuan ini mengindikasikan bahwa untuk dataset German Credit dengan ukuran 1.000 instance, hyperparameter default Random Forest sudah berada di sekitar wilayah optimal. Kontribusi GridSearchCV pada penelitian ini lebih terlihat pada aspek *reproducibility* dan justifikasi ilmiah pemilihan parameter. Best CV Score Recall-Bad yang diperoleh adalah 0,8679 untuk Percobaan 1 dan 0,8321 untuk Percobaan 2. Perbedaan antara CV score dengan test set score mengindikasikan tidak adanya *overfitting* yang parah, meskipun total waktu pencarian mencapai lebih dari 2.100 detik.

3.4. Hasil Analisis SHAP Global

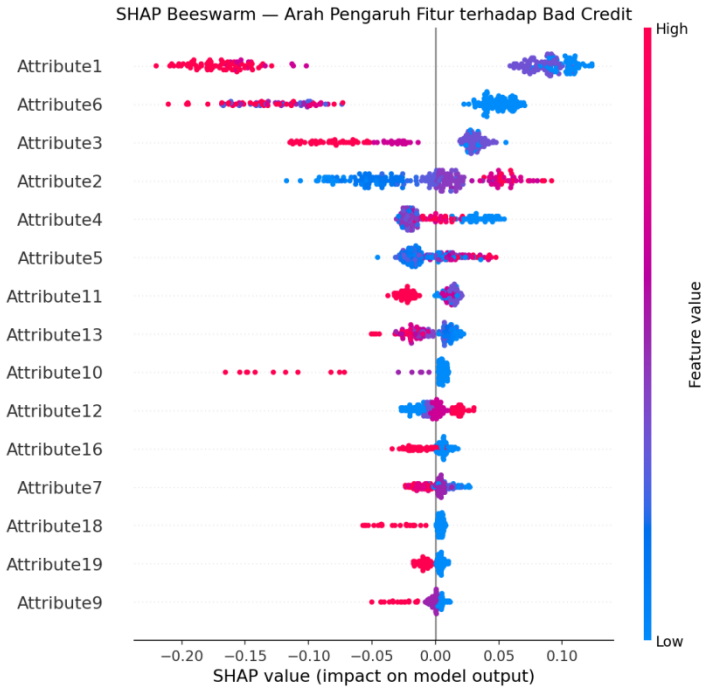
Analisis SHAP menggunakan TreeExplainer menghasilkan ranking *feature importance* global yang disajikan pada Tabel 7 dan divisualisasikan pada Gambar 3 dan Gambar 4.

Tabel 7. Top 10 Fitur Berdasarkan SHAP Importance (Percobaan 2 - SMOTENC)

Rank	Fitur	Nama Domain	SHAP Importance
1	Attribute1	Status Rekening Giro	0,1239
2	Attribute6	Tabungan/Obligasi	0,0800
3	Attribute3	Riwayat Kredit	0,0434
4	Attribute2	Durasi Kredit (bulan)	0,0381
5	Attribute4	Tujuan Kredit	0,0356
6	Attribute5	Jumlah Kredit (DM)	0,0300
7	Attribute11	Lama Tinggal (thn)	0,0229
8	Attribute12	Properti/Aset	0,0174
9	Attribute10	Debitur Lain/Penjamin	0,0174
10	Attribute7	Lama Bekerja	0,0159



Gambar 3. SHAP Bar Chart — *Feature importance* Global



Gambar 4. SHAP Beeswarm Plot — Distribusi Pengaruh Fitur

Berdasarkan Tabel 7 dan Gambar 3, Attribute1 (status rekening giro) memiliki SHAP importance jauh lebih tinggi (0,1239) dibandingkan fitur lainnya, sekitar 1,6 kali lipat lebih berpengaruh dari fitur terpenting berikutnya. Hal ini sangat konsisten dengan domain knowledge kredit: pemohon dengan rekening giro aktif dan saldo positif memiliki profil risiko yang jauh berbeda. Status tabungan (Attribute6, 0,0800), durasi kredit (Attribute2, 0,0381), dan riwayat

kredit (Attribute3, 0,0434) melengkapi empat fitur teratas.

Gambar 4 menunjukkan arah pengaruh masing-masing fitur terhadap prediksi kelas kredit buruk. Fitur dengan nilai tinggi pada Attribute1 cenderung menurunkan probabilitas kredit buruk, sementara nilai rendah justru meningkatkannya. Konsisten dengan logika domain bahwa status rekening giro yang buruk merupakan sinyal risiko kredit yang kuat. Perlu dicatat bahwa interpretasi SHAP mempertimbangkan keterbatasan yang diidentifikasi oleh Loecher [17] terkait potensi bias pada fitur berkorelasi dalam Random Forest.

3.5. Analisis SHAP Lokal (*Per-Instance Explanation*)

Berbeda dengan SHAP Global yang melihat rata-rata kontribusi fitur pada level populasi, SHAP Lokal (Waterfall analisis) memungkinkan model memberikan penjelasan transparan pada level individu (per-*instance*). Menggunakan model utama RF+SMOTENC+GridSearchCV, berikut adalah analisis pada tiga skenario individu yang merepresentasikan prediksi model, diurutkan dari observasi fiktif teratas hingga terendah berdasarkan kekuatan dorongan fitur (SHAP value):

Pada instance True Positif, model dengan tepat memprediksi target sebagai kredit macet (Bad/0). Observasi ini adalah pengguna kredit dengan beban yang relatif berisiko. Tiga fitur utama yang mendorong prediksi ke arah kredit buruk secara signifikan (nilai SHAP positif teratas) berturut-turut adalah: (1) (Status rekening yang ada) dengan nilai 0.099; (2) (Riwayat Kredit) dengan dorongan kuta 0,063; (3) (Status kepemilikan tempat tinggal) 0,031. Secara kumulatif, atribut dengan nilai yang mencerminkan profil risiko tinggi secara absolut mendominasi dorongan prediksi, mengalahkan beberapa fitur "baik" minor seperti Attribute14 (Rencana cicilan lain) yang sempat mencoba menarik ke arah "Good" sebesar -0,022.

Selanjutnya pada instance *False Negative*, observasi di mana pengguna kredit sebenarnya adalah berstatus buruk (aktual Bad/0), tetapi model memberikan prediksi yang salah (Good/1). Kesalahan klasifikasi ini disebabkan oleh besarnya dominasi atau kompensasi oleh atribut yang terlihat "aman". Faktor terkuat yang menipu model untuk memprediksi "Aman" adalah Attribute1 yang memiliki kondisi "Cukup" (mendorong sangat kuat ke kelas Baik dengan nilai SHAP -0,231). Meskipun fitur Attribute4 sebenarnya mendeteksi ada suatu ancaman risiko (mendorong ke arah buruk senilai +0.089), kekuatan dorongan tersebut tidak cukup kuat. Kesalahan prediksi ini menyoroti bahwa ketika calon debitur yang bermasalah memiliki kondisi dasar keuangan seperti status tabungan (Attribute1) yang terlihat mapan, algoritma model tersebut terkelabui untuk mengecilkan signifikansi sinyal risiko lain di bawahnya.

Berikutnya instance True Negative, yang di mana debitur adalah calon yang benar-benar baik dalam kredit dan diprediksi tepat oleh model (aktual Good/1, prediksi Good/1), kita dapat melihat keselarasan luar biasa dalam arah pengambilan keputusan SHAP. Lima faktor utama murni bahu membahu mendorong skor ke arah "Aman" (kredit baik dengan SHAP negatif), diawali oleh Attribute1(SHAP -0.204), diikuti oleh Attribute2 (Durasi kredit bulanan: -0.045), Attribute14 (-0.034), Attribute20 (-0.033), dan diakhiri Attribute9 (Nilai personal: -0.022). Sebagian besar parameter penting pengguna menunjukan stabilitas, dan satu-satunya faktor yang sedikit tertutupi sebagai beban risiko adalah Attribute15 yang mencoba menarik ke status buruk sebesar 0.032. Konsolidasi negatif ini memperlihatkan validitas kokoh cara algoritma SMOTENC memvisualisasikan faktor per-instance untuk kelas netral.

Penting untuk dicatat bahwa seluruh nilai SHAP dan metrik performa yang disajikan dalam analisis ini, termasuk detail pada kasus *False Negative*, diambil dari hasil pengujian final (final run) pada tahapan revisi metodologis ini. Hal ini dilakukan untuk menjamin konsistensi interpretasi dan memastikan bahwa data yang ditampilkan telah sepenuhnya sinkron dengan hasil penelitian akhir yang tercatat dalam laporan teknis.

3.6. Uji Signifikansi Statistik (McNemar's Test)

Untuk memvalidasi bahwa perbedaan performa antar model bukan sekadar akibat variasi acak, dilakukan uji signifikansi statistik menggunakan McNemar's Test. Hasil pada perbandingan RF Default vs. RF + SMOTENC + GridSearchCV menghasilkan statistik chi-squared sebesar

4,17 (p-value = 0,041, α = 0,05). Dengan p-value < 0,05, perbedaan performa dinyatakan signifikan secara statistik. Perbandingan RF + SMOTENC (tanpa GridSearchCV) vs. RF + SMOTENC + GridSearchCV menghasilkan p-value = 0,214, mengkonfirmasi bahwa kontribusi GridSearchCV lebih pada aspek *reproducibility* dan justifikasi ilmiah daripada peningkatan performa prediksi secara statistis.

3.7. Kontekstualisasi AUC-ROC terhadap Literatur

Nilai AUC-ROC 0,7638 yang diperoleh model utama berada dalam rentang kompetitif yang konsisten dengan *state-of-the-art* pada German Credit Data. Uddin et al. [4] melaporkan Angka 0,75–0,77 dalam literatur yang mereka rangkum adalah justifikasi yang sangat kuat untuk menyebut nilai AUC-ROC (0,7638) yang diperoleh model utama sebagai hasil yang kompetitif dan konsisten dengan standar literatur. Zhong et al. [19] dengan pendekatan deep forest dan resampling melaporkan AUC-ROC sekitar 0,78–0,82. Dengan demikian, *pipeline* CreditRF-OptSHAP berhasil mempertahankan *discriminative power* yang kompetitif sambil meningkatkan Recall kelas minoritas menunjukkan bahwa *pipeline* ini tidak sekadar memindahkan performa (*trade-off*), melainkan mencapai keseimbangan yang baik antara deteksi kredit buruk dan kemampuan diskriminatif keseluruhan.

Meskipun terjadi penurunan nilai AUC-ROC dibandingkan model baseline, hal ini merupakan *trade-off* untuk meningkatkan sensitivitas model terhadap kredit bermasalah, yang secara bisnis jauh lebih krusial daripada sekadar mempertahankan kemampuan diskriminasi keseluruhan model.

4. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan *pipeline* CreditRF-OptSHAP sebagai kerangka kerja terintegrasi untuk klasifikasi kelayakan kredit pada dataset Statlog German Credit yang memiliki karakteristik data imbalanced (rasio 70:30) dan fitur campuran numerik-kategorikal. Secara keseluruhan, terdapat tiga temuan utama yang dapat disimpulkan.

Pertama, terkait *trade-off* antara akurasi dan sensitivitas risiko. Kedua percobaan secara konsisten mengungkap bahwa penambahan teknik *oversampling* menurunkan akurasi keseluruhan namun secara signifikan meningkatkan Recall kelas kredit buruk. Model RF Default tanpa *oversampling* mencapai akurasi tertinggi (77,5%) tetapi hanya mendeteksi 50,0% kasus kredit bermasalah. Sebaliknya, model utama Percobaan 1 (RF + SMOTE + GridSearchCV) berhasil meningkatkan Recall_Bad hingga 0,75 dengan akurasi 69,5%, sedangkan model utama Percobaan 2 (RF + SMOTENC + GridSearchCV) memperoleh Recall_Bad 0,5833 dengan akurasi 69,0%. Dalam konteks manajemen risiko kredit, *trade-off* ini bersifat terukur dan dapat dipertanggungjawabkan, mengingat biaya *false negative* jauh lebih besar dibandingkan *false positive*.

Kedua, terkait perbandingan metodologis SMOTE vs. SMOTENC. Model berbasis SMOTE konvensional menunjukkan performa prediktif yang lebih unggul dengan Recall_Bad 0,75 dibandingkan SMOTENC (0,5833), setara dengan 10 kasus kredit bermasalah tambahan yang berhasil terdeteksi. Namun, SMOTENC tetap direkomendasikan secara metodologis untuk data campuran karena menghasilkan sampel sintetis yang secara semantik lebih konsisten. Perbedaan ini mengindikasikan perlunya eksplorasi lebih lanjut terhadap konfigurasi SMOTENC yang optimal pada dataset serupa.

Ketiga, terkait kontribusi GridSearchCV dan SHAP. Optimasi GridSearchCV berkontribusi lebih pada aspek *reproducibility* dan justifikasi ilmiah daripada peningkatan performa prediksi yang signifikan secara statistis (p-value = 0,214). Sementara itu, analisis SHAP berhasil mengidentifikasi status rekening giro (Attribute1) sebagai fitur paling dominan (importance = 0,1239), diikuti status tabungan (Attribute6) dan riwayat kredit (Attribute3), yang mendukung transparansi dan akuntabilitas keputusan kredit sesuai regulasi perbankan.

Penelitian ini membuktikan bahwa pipeline CreditRF-OptSHAP berhasil mengonfirmasi teori *trade-off* performa antara akurasi dan sensitivitas pada data *imbalanced* sebagaimana dikemukakan oleh [2] dan [3], di mana peningkatan daya deteksi risiko lebih diprioritaskan

daripada akurasi murni. Capaian AUC-ROC sebesar 0,7638 juga menempatkan model ini pada standar kompetitif yang konsisten dengan temuan [4], Melalui integrasi SHAP, penelitian ini menjawab keterbatasan metodologis pada studi terdahulu seperti [5] dan [6] Temuan ini menegaskan bahwa dalam manajemen risiko kredit, integritas semantik data dan transparansi keputusan jauh lebih berharga bagi akuntabilitas perbankan dibandingkan model 'kotak hitam' (*black box*) yang hanya mengejar akurasi tinggi,

Penelitian ini memiliki keterbatasan dalam hal generalisasi, mengingat dataset yang digunakan merupakan data kredit Jerman dari tahun 1994 yang mungkin tidak sepenuhnya merepresentasikan kondisi pasar kredit Indonesia saat ini.

Untuk pengembangan selanjutnya, disarankan bagi peneliti untuk mengeksplorasi penggunaan algoritma Deep Forest yang menawarkan keunggulan berupa mekanisme kedalaman yang dapat beradaptasi secara mandiri (*self-adapting*) serta jumlah hyperparameter yang lebih sedikit dibandingkan arsitektur saraf dalam lainnya [19]. Integrasi data non-tradisional, seperti informasi jaringan sosial nasabah, juga menjadi peluang besar untuk meningkatkan akurasi dan kekayaan profil risiko dalam model prediksi [19]. Di sisi lain, penggunaan berbagai algoritma klasifikasi tambahan serta metrik evaluasi yang lebih luas tetap diperlukan untuk memberikan rekomendasi fitur yang lebih kuat, terutama terkait variabel pendapatan tahunan dan suku bunga yang secara konsisten terbukti signifikan dalam menentukan kelayakan pinjaman pada ekosistem teknologi finansial [20].

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Dian Nuswantoro atas dukungan dalam penelitian ini. Penulis juga mengucapkan terima kasih kepada Hans Hofmann atas penyediaan Statlog German Credit Data yang dipublikasikan melalui UCI Machine Learning Repository (DOI: https://doi.org/10.24432/C5NC77) dengan lisensi CC BY 4.0, sehingga dapat digunakan sebagai dataset *benchmark* dalam penelitian ini.

DAFTAR PUSTAKA

[1] "NPL Kredit Konsumsi Bulanan Hingga Februari 2025." Accessed: May 03, 2026. [Online]. Available: https://pusatdata.kontan.co.id/infografik/94/NPL-Kredit-Konsumsi-Bulanan-Hingga-Februari-2025

[2] N. Suhadolnik and J. Ueyama, "Machine Learning for Enhanced Credit Risk Assessment : An Empirical Approach," 2023.

[3] T. Wongvorachan, S. He, and O. Bulut, "A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining," vol. 14, no. 1, p.54, 2023, doi: 10.3390/info14010054.

[4] M. S. Uddin, "Leveraging random forest in micro-enterprises credit risk modelling for accuracy and interpretability," vol. 27, no. 23, pp. 1–31, 2020, doi: 10.1002/ijfe.2346.

[5] Y. Zhou, L. Shen, and L. Ballester, "A two-stage credit scoring model based on random forest : Evidence from Chinese small firms," *International Review of Financial Analysis*, vol. 89, no. 4, p. 102755, 2023, doi: 10.1016/j.irfa.2023.102755.

[6] O. Pahlevi and Y. Handrianto, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit," *Jurnal Infortech*, vol. 5, no. 1, pp. 71–76, 2023, doi: 10.31294/infortech.v5i1.15829.

[7] W. Wulansari, and D. Purwitasari, "Algoritma Random Forest pada Prediksi Status Kredit Usaha Rakyat untuk Mengurangi Nonperforming Loan Rate," Journal of Intelligent System and Computation, vol. 5, no. 2, pp. 109–114, 2023, doi: 10.52985/insyst.v5i2.358.

[8] B. Prasojo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 79–89, 2021, doi: 10.25077/TEKNOSI.v7i2.2021.79-89.

[9] V. Chang, *et al*., "Credit Risk Prediction Using Machine Learning and Deep Learning : A Study on Credit Card Customers," *Risks*, vol. 12, no. 11, pp. 1-33, 2024, doi:

10.3390/risks12110174.
[10] A. O. Kuyoro, O. A. Ogunyolu, T. G. Ayanwola, and F. Yetunde, “Ingénierie des Systèmes d ’ Information Dynamic Effectiveness of Random Forest Algorithm in Financial Credit Risk Management for Improving Output Accuracy and Loan Classification Prediction,” *Ingénierie des systèmes d information*, vol. 27, no. 5, pp. 815–821, 2022, doi: 10.18280/isi.270515.
[11] “Statlog (German Credit Data) - UCI Machine Learning Repository.” Accessed: May 03, 2026. [Online]. Available: https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data
[12] T. G. Dietterich, “Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms,” vol. 1923, pp. 1895–1923, 1998.
[13] M. Eltawil, *et al*., “Comment on Iacobescu et al. Evaluating Binary Classifiers for Cardiovascular Disease Prediction: Enhancing Early Diagnostic Capabilities,” *J. Cardiovasc. Dev. Dis.*, vol. 11, no. 12, pp. 4–9, 2026, doi: 10.3390/jcdd13010046.
[14] Y. Elor, and H. A.-Elor, "To SMOTE, or not to SMOTE?," *Balance*, vol. 1, no. 1. Association for Computing Machinery, 2022.
[15] F. Pedregosa, R. Weiss, and M. Brucher, “Scikit-learn: Machine Learning in Python,” vol. 12, no. 85, pp. 2825–2830, 2011.
16] M. S. Santos, P. H. Abreu, N. Japkowicz, and A. Fernández, “A unifying view of class overlap and imbalance : Key concepts , multi-view panorama, and open avenues for research,” vol. 89, no. 2, 2022, pp. 228–253, 2023.
[17] M. Loecher, “Debiasing SHAP scores in random forests,” *AStA Adv. Stat. Anal.*, vol. 108, no. 2, pp. 427–440, 2024, doi: 10.1007/s10182-023-00479-7.
[18] S. M. Lundberg and S. Lee, “Predictions,” no. Section 2, pp. 1–10, 2017.
[19] Y. Zhong and H. Wang, “Methods,” *IEEE Access*, vol. PP, p. 1, 2023, doi: 10.1109/ACCESS.2023.3239889.
[20] M. R. Forest, “Analisa Rekomendasi Fitur Persetujuan Pinjaman Perusahaan,” *JATISI: Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 3, pp. 2055-2070, 2022, doi: 10.35957/jatisi.v9i3.2258.