

Perbandingan Prediksi *Loan Default* Menggunakan Pemodelan Algoritma *Random Forest* dan *XGBoost* untuk Mitigasi Risiko Kredit Bank (Studi Kasus: BTN Syariah)

Gigih Prakoso Wigantiyoko¹, R. Mohamad Atok²

^{1,2} Program Magister Manajemen Teknologi, Sekolah Interdisiplin Manajemen dan Teknologi (SIMT), Institut Teknologi Sepuluh Nopember
Jl. Cokroaminoto No. 12A, Surabaya, Jawa Timur 60264
Email: gigihprakosow@gmail.com

ABSTRAK

Peningkatan penyaluran pembiayaan syariah pascapandemi meningkatkan risiko gagal bayar yang mengancam stabilitas bank. Bank BTN Syariah Cabang X mencatat *Non Performing Financing* (NPF) sebesar 15,27% pada 2019, turun ke 1,47% pada 2022, namun meningkat kembali pada 2023-2024. Penelitian ini membangun dan membandingkan model prediksi gagal bayar pembiayaan konsumen menggunakan algoritma *Random Forest* dengan *Synthetic Minority Over-sampling Technique* (SMOTE) dan *Extreme Gradient Boosting* (XGBoost) dengan *Weighted Class* sebagai fondasi *Early Warning System* (EWS) untuk monitoring nasabah eksisting secara periodik. Data yang digunakan adalah 14.374 nasabah pembiayaan konsumen aktif per 31 Desember 2024 dengan rasio ketidakseimbangan kelas sebesar 77,5:1. Model dievaluasi menggunakan *recall*, *precision*, *F1-Score*, *ROC-AUC*, dan *PR-AUC* dengan penekanan pada *recall* sebagai metrik prioritas. Hasil menunjukkan bahwa *Random Forest* dengan SMOTE merupakan model terbaik dengan *recall* kelas *default* = 0,6757 dan *ROC-AUC* = 0,9368, mendeteksi 25 dari 37 nasabah gagal bayar pada data testing. XGBoost menghasilkan *F1-Score* = 0,4878 dan *PR-AUC* = 0,4849 lebih tinggi namun *recall* lebih rendah (0,5405). Analisis *feature importance* menggunakan MDI dan SHAP mengungkapkan bahwa Saldo Tersedia dan Rasio Saldo/Angsuran merupakan prediktor dominan, membuktikan likuiditas *real time* sebagai *leading indicator* yang lebih kuat dari riwayat tunggakan.

Kata kunci: Gagal Bayar, *Non Performing Financing*, *Random Forest*, XGBoost, *Early Warning System*, Risiko Kredit

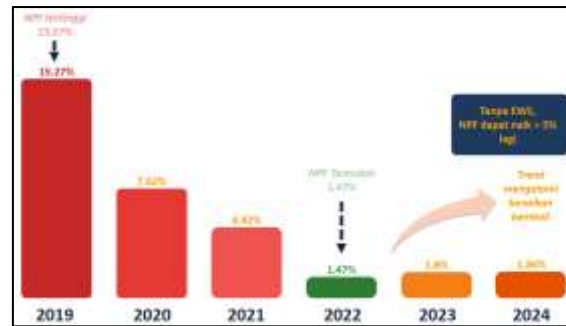
ABSTRACT

Post-pandemic expansion of Islamic banking financing has elevated default risk, threatening bank stability. Bank BTN Syariah Branch X recorded NPF of 15.27% in 2019, reduced to 1.47% by 2022, before rising again in 2023-2024. This study develops and compares consumer financing default prediction models using *Random Forest* with *Synthetic Minority Over-sampling Technique* (SMOTE) and *Extreme Gradient Boosting* (XGBoost) with *Weighted Class* as the analytical foundation of a periodic *Early Warning System* (EWS) for monitoring existing borrowers. The dataset comprises 14,374 active consumer financing customers as of 31 December 2024 with a class imbalance ratio of 77.5:1. Models are evaluated using *recall*, *precision*, *F1-Score*, *ROC-AUC*, and *PR-AUC* with *recall* prioritized as the primary metric. Results show that *Random Forest* with SMOTE is the superior model achieving default-class *recall* of 0.6757 and *ROC-AUC* of 0.9368, detecting 25 of 37 defaulting customers in the test set. XGBoost yields higher *F1-Score* (0.4878) and *PR-AUC* (0.4849) but lower *recall* (0.5405). MDI and SHAP *feature importance* analysis identifies Available Balance and Balance-to-Installment Ratio as dominant predictors, confirming real-time liquidity as a stronger leading indicator than delinquency history.

Keywords: Loan Default, *Non Performing Financing*, *Random Forest*, XGBoost, *Early Warning System*, Credit Risk

Pendahuluan

Non Performing Financing (NPF) merupakan indikator utama kualitas pembiayaan yang secara langsung memengaruhi profitabilitas dan kestabilan bank[1]. Bank BTN Syariah sebagai penyalur utama Kredit Pemilikan Rumah (KPR) Subsidi mencatat pola NPF yang berfluktuasi: 15,27% pada 2019, berhasil ditekan ke 1,47% pada 2022, namun meningkat kembali pada 2023-2024. Pola bouncing pada **gambar 1** ini mengindikasikan bahwa pendekatan mitigasi risiko konvensional yang bersifat reaktif belum optimal, khususnya di Bank BTN Syariah Cabang X yang menjadi obyek penelitian ini[2].



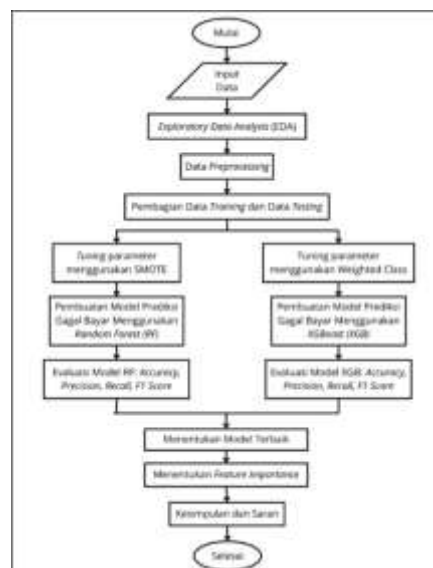
Gambar 1. Prosentase NPF BTN syariah cabang “X”

Prediksi gagal bayar berbasis machine learning telah terbukti lebih efektif dibandingkan metode statistik tradisional dalam mendeteksi risiko kredit secara dini[3]. Algoritma ensemble seperti Random Forest dan XGBoost menunjukkan performa superior pada data kredit yang kompleks dan tidak seimbang[4], [5]. Data kredit secara alamiah bersifat tidak seimbang, dan studi eksperimental menunjukkan bahwa pemilihan algoritma serta strategi penanganan imbalance secara tepat berpengaruh signifikan terhadap kinerja klasifikasi risiko kredit[6]. Namun demikian, sebagian besar penelitian yang ada berfokus pada credit scoring untuk calon nasabah baru, bukan pada sistem peringatan dini (Early Warning System/EWS) untuk nasabah eksisting[7]. Tinjauan sistematis terbaru menegaskan bahwa penanganan imbalanced learning masih menjadi tantangan utama dalam prediksi gagal bayar kredit[8]. Lebih spesifik lagi, studi yang secara eksplisit menangani rasio ketidakseimbangan kelas yang sangat ekstrem (77,5:1) sebagaimana ditemukan pada portofolio pembiayaan konsumen eksisting masih jarang dibahas dalam literatur, sehingga menyisakan celah metodologis antara teknik penanganan imbalanced class yang tersedia dengan kondisi data riil di perbankan syariah.

Penelitian ini mengisi kesenjangan tersebut dengan membangun model prediksi gagal bayar berbasis data historis portofolio pembiayaan konsumen Bank BTN Syariah sebagai fondasi EWS periodik untuk monitoring nasabah eksisting. Dua algoritma dibandingkan dengan pendekatan penanganan imbalanced class yang berbeda: Random Forest dengan SMOTE [9] dan XGBoost dengan Weighted Class[10]. Penelitian ini memberikan tiga kontribusi orisinal: (1) demonstrasi empiris pada data pembiayaan syariah Indonesia; (2) temuan bahwa variabel likuiditas real-time lebih prediktif dari riwayat tunggakan; dan (3) konstruksi fitur berbasis domain knowledge perbankan syariah yang belum ada dalam literatur sebelumnya. Secara praktis, model ini memberikan manfaat langsung bagi bank berupa deteksi dini nasabah berpotensi gagal bayar sebelum tunggakan terjadi, mitigasi risiko NPF melalui intervensi preventif oleh relationship manager, serta pemanfaatan variabel likuiditas real-time (Saldo Tersedia dan Rasio Saldo/Angsuran) sebagai leading indicator yang dapat dipantau secara periodik tanpa menunggu keterlambatan pembayaran terjadi.

Metode Penelitian

Alur Penelitian



Gambar 2. Alur penelitian

Gambar 2 diatas menunjukkan alur penelitian yang dimulai dengan melakukan *Exploratory Data Analysis* (EDA), kemudian dilanjutkan dengan *data preprocessing*, pemodelan *Random Forest* serta *XGBoost*, dan menentukan fitur paling berpengaruh terhadap *loan default*.

Sumber Data

Data penelitian bersumber dari Bank BTN Syariah Cabang X, terdiri dari 14.374 nasabah pembiayaan konsumen aktif per 31 Desember 2024 dengan 30 variabel numerik dan kategorikal. Label target default didefinisikan menggunakan dua rule: (1) kolektibilitas ≥ 3 sesuai POJK No. 2/POJK.03/2022, dan (2) tunggakan > 0 dan periode menunggak > 3 bulan sesuai SE OJK No. 34/SEOJK.03/2021. Dari total 14.374 nasabah, 183 nasabah (1,27%) tergolong default dan 14.191 (98,73%) non-default, menghasilkan rasio ketidakseimbangan kelas yang sangat ekstrem sebesar 77,5:1.

Feature Engineering

Enam fitur baru dikonstruksi berdasarkan domain knowledge perbankan syariah. Kelompok pertama [5B] mengekstrak dimensi temporal dari variabel tanggal: *Usia_Pembiayaan_Bln* (usia pembiayaan sejak akad), *Sisa_Tenor_Bln* (sisa masa jatuh tempo), dan *Total_Tenor_Bln* (total jangka waktu akad). Kelompok kedua [5D] mengonstruksi rasio berbasis saldo blokir: *Rasio_Blokir_Angsuran*, *Flag_Blokir_Habis* (biner: 1 jika saldo blokir < 1 kali angsuran), dan *Rasio_Blokir_Outstanding*. Fitur-fitur rasio saldo ini merepresentasikan kondisi likuiditas nasabah sebagai leading indicator yang secara kausal mendahului kejadian gagal bayar [11]. Secara praktis, ketiga fitur temporal pada kelompok pertama membantu mengidentifikasi fase siklus hidup pembiayaan yang paling rentan terhadap gagal bayar; nasabah pada tahun-tahun awal akad dengan sisa tenor yang masih panjang cenderung memiliki eksposur risiko lebih tinggi karena belum terbentuk rekam jejak pembayaran yang panjang. Secara operasional, *Rasio_Blokir_Angsuran* dan *Flag_Blokir_Habis* memberikan sinyal dini kepada relationship manager mengenai kecukupan dana cadangan nasabah untuk menutup kewajiban angsuran berikutnya, sementara *Rasio_Blokir_Outstanding* menggambarkan proporsi eksposur kredit yang telah ter-cover oleh saldo blokir sebagai bentuk mitigasi risiko bawaan (built-in) pada struktur pembiayaan.

Data Preprocessing untuk Data Skewed

Statistik deskriptif (lihat bagian Statistik Deskriptif) menunjukkan karakteristik distribusi positively skewed dengan outlier ekstrem pada variabel numerik seperti *Outstanding Pokok*. Berbeda dengan algoritma berbasis jarak seperti k-NN atau regresi logistik yang sensitif terhadap skala fitur, *Random Forest* dan *XGBoost* merupakan algoritma berbasis pohon keputusan yang invarian terhadap transformasi monoton pada fitur numerik, karena proses pemilihan split point hanya bergantung pada urutan (ranking) nilai, bukan skala absolutnya[12]. Dengan demikian, standarisasi atau normalisasi fitur tidak diperlukan secara wajib agar model dapat mencapai kinerja optimal. Outlier ekstrem pada variabel seperti *Outstanding Pokok* tetap dipertahankan tanpa capping atau transformasi karena nilai tersebut merepresentasikan eksposur kredit riil yang relevan secara bisnis; penghapusan atau transformasi outlier berisiko menghilangkan sinyal penting terkait nasabah bernilai tinggi. Tahapan pembersihan data juga mencakup penanganan missing value melalui strategi imputasi yang merupakan langkah standar pada data kredit.[13] Meskipun demikian, untuk penelitian lanjutan yang melibatkan algoritma berbasis jarak atau linear, penerapan robust scaling atau log transformation pada variabel dengan skewness tinggi direkomendasikan sebagai langkah preprocessing tambahan sebelum pemodelan.

Model Random Forest SMOTE

Random Forest dipilih karena kemampuannya menangani data tidak seimbang dan fitur tidak relevan secara robust[4]. Ketidakseimbangan kelas ditangani menggunakan *SMOTE* yang menghasilkan sampel sintetis kelas minoritas melalui interpolasi k-nearest neighbors[9]. *SMOTE* ditempatkan di dalam *ImbPipeline*, bukan di luar, untuk mencegah data leakage selama cross-validation[14]. Optimasi hyperparameter menggunakan *HalvingRandomSearchCV* berdasarkan algoritma successive halving yang terbukti 3 sampai dengan 5 kali lebih efisien dari *GridSearchCV*[15], dengan F1-Score kelas minoritas sebagai scoring metric karena lebih informatif dari accuracy pada kondisi imbalanced[16].

Model XGBoost Weighted Class

XGBoost menangani ketidakseimbangan kelas secara internal melalui parameter *scale_pos_weight* yang memodifikasi gradient dan hessian kelas positif selama boosting sekuensial[10]. Nilai optimal *scale_pos_weight* dihitung dari rasio kelas negatif terhadap positif pada data training:

$$\text{scale_pos_weight} = \text{n_neg} / \text{n_pos} = 9.683 / 146 = 66,32$$

Nilai ini berarti model memberikan bobot 66 kali lebih besar pada kesalahan klasifikasi kelas default. Penggunaan *tree_method='hist'* mempertahankan akurasi identik dengan metode exact namun 5 sampai dengan 10 kali lebih efisien secara komputasional[17]. Regularisasi L1 (*reg_alpha*) dan L2 (*reg_lambda*) dioptimalkan

bersama untuk mencegah overfitting[18]. Tuning menggunakan RandomizedSearchCV (n_iter=20) yang efisien untuk ruang parameter yang luas[19].

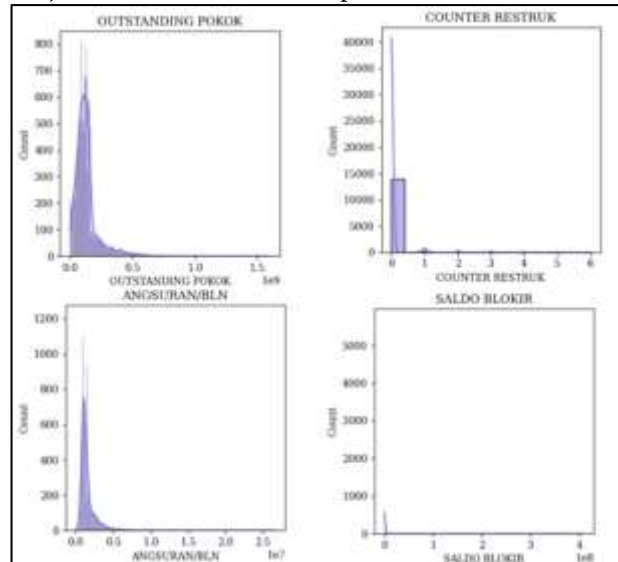
Evaluasi Model

Evaluasi menggunakan lima metrik: recall, precision, F1-Score, ROC-AUC, dan PR-AUC. Recall kelas default diprioritaskan sebagai metrik utama karena biaya False Negative (nasabah default tidak terdeteksi) secara finansial jauh lebih besar dari biaya False Positive dalam konteks EWS perbankan[20]]. PR-AUC digunakan sebagai metrik diskriminasi utama karena lebih informatif dari ROC-AUC pada kondisi imbalanced ekstrem. ROC-AUC dapat memberikan gambaran yang terlalu optimistis karena tidak melibatkan True Negative[21]. Feature importance divalidasi menggunakan MDI (Mean Decrease Impurity) dan SHAP (SHapley Additive exPlanations) secara paralel untuk mendeteksi potensi bias kardinalitas pada MDI, sejalan dengan pendekatan explainable machine learning yang direkomendasikan untuk prediksi gagal bayar kredit[22], [23].

Hasil Dan Pembahasan

Statistik Deskriptif

Dataset menunjukkan karakteristik distribusi positively skewed pada **gambar 3** untuk empat variabel numerik. Outstanding Pokok memiliki mean Rp224.723.531 dengan standar deviasi Rp12,45 miliar menunjukkan kesenjangan yang mengindikasikan distribusi log normal dengan outlier ekstrem hingga Rp1,38 triliun. Variabel perilaku seperti Count Tunggakan Bulan (mean=0,20; median=0; max=15) dan Counter Restruk (mean=0,06; median=0; max=6) menunjukkan pola zero-inflated yang konsisten dengan dominasi nasabah kolektibilitas 1 (lancar) sebesar 93,55% dari total portofolio.



Gambar 3. Distribusi data untuk variabel outstanding pokok, counter restruk, angsuran/bln, dan saldo blokir

Hasil Model Random Forest SMOTE

Proses HalvingRandomSearchCV berlangsung dalam 4 iterasi (305 total fits, 32,7 menit) menghasilkan Best CV F1-Score = 0,3161 dengan parameter optimal: n_estimators=400, max_depth=10, max_features=0,3, smote__sampling_strategy=1,0, smote__k_neighbors=3, dan class_weight='balanced'. **Tabel 1** menyajikan confusion matrix dan metrik evaluasi model RF pada data testing.

Tabel 1. Confusion matrix dan metrik evaluasi model Random Forest SMOTE

Metrik	Kelas 0 (Aman)	Kelas 1 (Default)	Keterangan
Recall	0,9620	0,6757	★ Prioritas EWS
Precision	0,9949	0,2137	Trade-off imbalanced
F1-Score	0,9782	0,3247	Macro avg: 0,6514
ROC-AUC	—	0,9368	Diskriminasi sangat baik
PR-AUC	—	0,3923	Informatif untuk imbalanced
TP/FN	—	25 / 12	25 default terdeteksi

Hasil Model XGBoost Weighted Class

RandomizedSearchCV (20 kandidat, 13,3 menit) pada **Tabel 2** menghasilkan Best CV F1-Score = 0,4735 dengan parameter optimal: n_estimators=150, max_depth=7, learning_rate=0,15, subsample=0,6, colsample_bytree=0,6, min_child_weight=5, reg_alpha=0,1, reg_lambda=2,0, dan scale_pos_weight=66,32. **Tabel 2** menyajikan metrik evaluasi XGBoost.

Tabel 2. Confusion matrix dan metrik evaluasi model XGBoost Weighted Class

Metrik	Kelas 0 (Aman)	Kelas 1 (Default)	Keterangan
Recall	0,9897	0,5405	Lebih rendah dari RF
Precision	0,9930	0,4444	★ Presisi lebih baik
F1-Score	0,9913	0,4878	★ Macro avg: 0,7396
ROC-AUC	—	0,9321	Sangat baik
PR-AUC	—	0,4849	★ Lebih baik dari RF
TP/FN	—	20 / 17	17 default terlewat

Perbandingan dan Pemilihan Model Terbaik

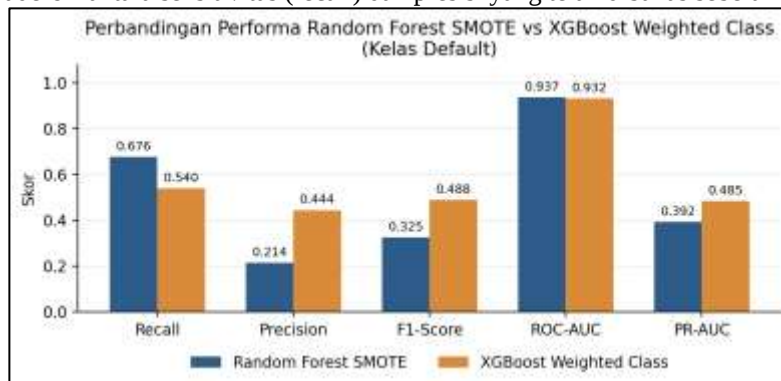
Berdasarkan perbandingan komprehensif pada **Tabel 3**, Random Forest SMOTE dipilih sebagai model terbaik berdasarkan tiga pertimbangan utama. Pertama, recall kelas default tertinggi (0,6757 vs 0,5405), mendeteksi 25 dari 37 nasabah default (False Negative hanya 12 vs 17 pada XGBoost). Dalam konteks EWS perbankan, biaya False Negative (NPF tersembunyi) secara finansial jauh lebih besar dari biaya False Positive (verifikasi relationship manager)[20]. Kedua, ROC-AUC tertinggi (0,9368 vs 0,9321) memberikan fondasi yang lebih kuat untuk threshold optimization. Ketiga, profil portofolio yang didominasi KPR Non Subsidi (70,36%) mengindikasikan toleransi terhadap false alarm lebih tinggi dibandingkan dampak NPF tidak terdeteksi.

Tabel 3. Perbandingan hasil pemodelan Random Forest SMOTE dan XGBoost Weighted Class

Metrik	RF + SMOTE	XGBoost + WC	Unggul
Recall — Deteksi Default	0,6757	0,5405	RF
False Negative (FN)	12	17	RF
False Positive (FP)	92	25	XGB
Precision	0,2137	0,4444	XGB
F1-Score	0,3247	0,4878	XGB
ROC-AUC	0,9368	0,9321	RF
PR-AUC	0,3923	0,4849	XGB

XGBoost Weighted Class menunjukkan keunggulan pada F1-Score (0,4878 vs 0,3247), PR-AUC (0,4849 vs 0,3923), precision (0,4444 vs 0,2137), dan efisiensi komputasi (13,3 vs 32,7 menit). Model ini direkomendasikan sebagai alternatif ketika manajemen memprioritaskan presisi atas sensitivitas, misalnya ketika kapasitas tim relationship manager terbatas.

Perbandingan visual kelima metrik evaluasi antara kedua model disajikan pada Gambar 4, yang menegaskan pola trade-off antara sensitivitas (recall) dan presisi yang telah dibahas sebelumnya.



Gambar . Perbandingan performa Random Forest SMOTE vs XGBoost Weighted Class pada lima metrik evaluasi kelas default

Analisis Biaya-Manfaat (Cost-Benefit Analysis)

Untuk mengkuantifikasi implikasi finansial dari perbedaan recall antar model, dilakukan estimasi biaya-manfaat sederhana menggunakan rata-rata Outstanding Pokok (Rp224.723.531 per nasabah) sebagai proksi eksposur kredit per False Negative. Model Random Forest SMOTE melewati 12 nasabah default (False Negative), sedangkan XGBoost Weighted Class melewati 17 nasabah, selisih 5 nasabah. Dengan asumsi eksposur rata-rata tersebut, potensi eksposur risiko yang tidak terdeteksi secara ilustratif mencapai sekitar Rp2,70 miliar pada Random Forest SMOTE ($12 \times \text{Rp}224.723.531$) dan sekitar Rp3,82 miliar pada XGBoost Weighted Class ($17 \times \text{Rp}224.723.531$), atau selisih potensi eksposur sekitar Rp1,12 miliar yang berpotensi dihindari dengan memilih Random Forest SMOTE. Sebaliknya, biaya False Positive pada EWS relatif rendah karena hanya berupa biaya operasional verifikasi oleh relationship manager, jauh lebih kecil dibandingkan potensi kerugian NPF yang tidak terdeteksi. Perhitungan ini bersifat ilustratif (order-of-magnitude) karena belum memperhitungkan recovery rate agunan, biaya restrukturisasi, dan time value of money; penelitian lanjutan disarankan membangun cost matrix yang lebih presisi dengan melibatkan data recovery historis dan biaya operasional aktual bank.

Optimasi Threshold Probabilitas

Model Random Forest SMOTE menghasilkan ROC-AUC sebesar 0,9368, mengindikasikan kemampuan diskriminasi antar kelas yang sangat baik pada berbagai ambang batas probabilitas, sehingga optimasi threshold berpotensi meningkatkan recall lebih lanjut tanpa penurunan discriminative power yang signifikan. Threshold default sebesar 0,5 yang digunakan pada evaluasi Tabel 1 dan Tabel 2 belum tentu optimal untuk konteks EWS yang memprioritaskan recall. Salah satu pendekatan yang direkomendasikan adalah optimasi menggunakan Youden's J statistic ($J = \text{Sensitivitas} + \text{Spesifisitas} - 1$), yang mengidentifikasi titik pada kurva ROC dengan jarak vertikal maksimum terhadap garis diagonal random classifier, merepresentasikan keseimbangan optimal antara true positive rate dan false positive rate. Alternatif lain adalah pendekatan cost-sensitive threshold yang secara eksplisit memasukkan rasio biaya False Negative terhadap False Positive (lihat bagian Analisis Biaya-Manfaat) ke dalam penentuan titik potong optimal, mengingat kedua jenis kesalahan memiliki konsekuensi finansial yang sangat asimetris pada konteks EWS perbankan. Penerapan kedua pendekatan ini pada probabilitas prediksi Random Forest SMOTE direkomendasikan sebagai langkah lanjutan sebelum implementasi model pada sistem produksi, guna memaksimalkan deteksi nasabah berpotensi default tanpa mengorbankan efisiensi operasional tim relationship manager secara berlebihan.

Feature Importance

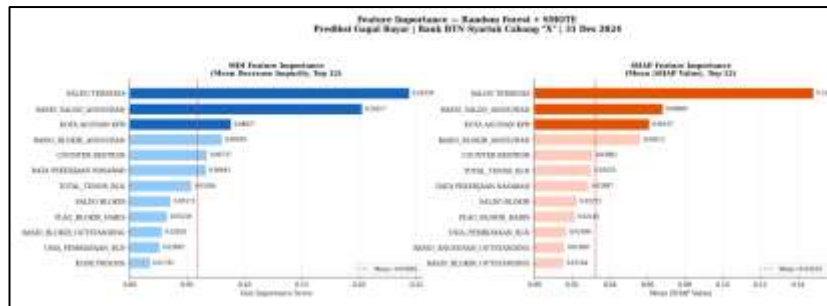
Analisis feature importance menggunakan MDI dan SHAP pada model Random Forest SMOTE menghasilkan ranking yang konsisten untuk seluruh 10 fitur teratas pada **Tabel 4**, mengkonfirmasi robustness hasil. **Tabel 4** menyajikan lima fitur teratas beserta skor MDI dan SHAP.

Tabel 4. Feature importance model Random Forest SMOTE

Rank	Fitur	MDI	SHAP
1	Saldo Tersedia	0,244	0,148
2	Rasio Saldo/Angsuran	0,203	0,068
3	Kota Agunan KPR	0,088	0,062
4	Rasio Blokir/Angsuran	0,080	0,056
5	Counter Restruk	0,067	0,031

Temuan orisinal penelitian ini adalah dominasi variabel likuiditas pada **Gambar 5** (Saldo Tersedia MDI=0,244; Rasio Saldo/Angsuran MDI=0,203) sebagai prediktor utama berbeda dari ekspektasi umum bahwa variabel tunggakan historis akan mendominasi. SHAP Beeswarm Plot mengkonfirmasi arah hubungan kausal yang konsisten dimana nilai saldo rendah (titik biru) secara konsisten mendorong prediksi default, membuktikan bahwa kondisi likuiditas real-time merupakan leading indicator yang mendahului keterlambatan pembayaran secara kausal[22].

Bagi manajer risiko, temuan ini memiliki implikasi operasional langsung: pemantauan EWS sebaiknya diprioritaskan pada saldo rekening dan rasio saldo terhadap angsuran secara berkala (misalnya bulanan), bukan semata-mata menunggu terjadinya tunggakan historis. Nasabah dengan Saldo Tersedia rendah dan Rasio Saldo/Angsuran di bawah ambang batas tertentu dapat dimasukkan ke dalam watchlist prioritas tinggi untuk intervensi dini oleh relationship manager, misalnya melalui restrukturisasi preventif atau edukasi pengelolaan likuiditas, sebelum kondisi berkembang menjadi tunggakan aktual.



Gambar 5. Feature importance model Random Forest SMOTE

Implikasi Manajerial dan Operasional

Hasil penelitian ini memberikan beberapa implikasi operasional bagi Bank BTN Syariah Cabang X. Pertama, tim relationship manager dapat memanfaatkan output model sebagai watchlist bulanan yang memprioritaskan nasabah dengan probabilitas prediksi tertinggi untuk dikunjungi atau dihubungi terlebih dahulu, sehingga sumber daya verifikasi dapat dialokasikan secara efisien. Kedua, strategi threshold EWS dapat dibedakan berdasarkan segmen produk; mengingat produk MR1/KPR Non Subsidi mencatat default rate tertinggi (2,84%), threshold yang lebih rendah (lebih sensitif) dapat diterapkan pada segmen ini untuk meningkatkan recall, sementara segmen dengan risiko lebih rendah seperti MMQ/Non KPR dapat menggunakan threshold standar untuk menjaga efisiensi operasional. Ketiga, prioritas intervensi nasabah dapat disusun berdasarkan kombinasi skor probabilitas model dan nilai Outstanding Pokok, sehingga nasabah dengan eksposur besar dan probabilitas default tinggi memperoleh prioritas penanganan tertinggi, sejalan dengan prinsip cost-benefit yang dibahas pada bagian sebelumnya.

Profil Nasabah Berpotensi Gagal Bayar

Gambar 6. Menunjukkan segmentasi jenis produk pembiayaan yang ada di BTN Syariah. Segmentasi berdasarkan kode produk menunjukkan default rate tertinggi pada produk MR1/KPR Non Subsidi (2,84%), memiliki nilai hampir dua kali rata-rata portofolio (1,49%). Sebaliknya, produk MMQ/Non KPR mencatat default rate terendah (0,36%), mengindikasikan bahwa segmen masyarakat berpenghasilan menengah sampai dengan tinggi memiliki risiko gagal bayar struktural yang lebih tinggi.



Gambar 6. Segmentasi produk pembiayaan dengan tingkat default rate tinggi

Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu menjadi perhatian dalam interpretasi hasil. Pertama, data yang digunakan hanya bersumber dari satu kantor cabang (Bank BTN Syariah Cabang X) dengan portofolio yang didominasi produk konsumen, sehingga karakteristik nasabah dan pola risiko yang teridentifikasi berpotensi tidak sepenuhnya merepresentasikan portofolio nasional Bank BTN Syariah secara keseluruhan; generalisasi model ke cabang atau segmen produk lain perlu dilakukan secara hati-hati dan idealnya divalidasi ulang menggunakan data multi-cabang. Kedua, model dilatih dan diuji menggunakan data historis yang ditarik pada satu titik waktu (snapshot per 31 Desember 2024), sehingga ketahanan model terhadap pergeseran pola risiko dari waktu ke waktu (concept drift) belum teruji. Pengujian out-of-time validation menggunakan data pada kuartal atau tahun yang berbeda diperlukan untuk memastikan performa model tetap stabil terhadap perubahan kondisi makroekonomi dan perilaku nasabah antar periode sebelum model diimplementasikan secara penuh pada sistem produksi.

Simpulan

Penelitian ini berhasil membangun dan membandingkan dua model prediksi gagal bayar pembiayaan konsumen berbasis machine learning pada data Bank BTN Syariah Cabang X. Model Random Forest dengan SMOTE terpilih sebagai model terbaik berdasarkan recall kelas default tertinggi (0,6757) dan ROC-AUC = 0,9368, mendeteksi 25 dari 37 nasabah default pada data testing. Model XGBoost dengan Weighted Class menghasilkan F1-Score (0,4878) dan PR-AUC (0,4849) yang lebih tinggi namun recall lebih rendah (0,5405) dengan keunggulan efisiensi komputasi 2,5 kali lebih cepat.

Temuan orisinal penelitian adalah dominasi variabel likuiditas real-time yaitu: Saldo Tersedia (MDI=0,244) dan Rasio Saldo/Angsuran (MDI=0,203) sebagai prediktor utama gagal bayar, mengkonfirmasi bahwa kondisi keuangan nasabah saat ini lebih prediktif dari riwayat keterlambatan masa lalu. Temuan ini mendukung reorientasi EWS dari monitoring lagging indicator ke leading indicator berbasis likuiditas.

Penelitian lanjutan direkomendasikan untuk mengeksplorasi: (1) validasi time-series multi-periode dengan out-of-time validation; (2) integrasi variabel makroekonomi eksternal; (3) eksplorasi algoritma LightGBM dan CatBoost[24] serta stacking ensemble; (4) implementasi SHAP local explanations untuk interpretabilitas individual; dan (5) replikasi pada data portofolio nasional untuk meningkatkan generalisabilitas model.

Daftar Pustaka

- [1] A. Yuwannita, R. Mulyany, and H. Fahlevi, "What Causes Non-Performance Financing? Insights From Islamic Commercial Banks in Indonesia and Malaysia," *Muqtasid J. Ekon. dan Perbank. Syariah*, vol. 13, no. 1, pp. 77–94, Oct. 2022, doi: 10.18326/muqtasid.v13i1.77-94.
- [2] S. Anwar and A. . H. Ali, "ANNs-Based Early Warning System for Indonesian Islamic Banks," *Bul. Ekon. Monet. dan Perbank.*, vol. 20, no. 3, pp. 325–342, Jan. 2018, doi: 10.21098/bemp.v20i3.856.
- [3] G. Chironna and G. Orlando, "Predicting bank defaults in Italy: A comparative analysis of conventional and machine learning approaches," *Econ. Anal. Policy*, vol. 89, pp. 788–833, Jan. 2026, doi: 10.1016/j.eap.2025.12.002.
- [4] L. Zhu, D. Qiu, D. Ergu, C. Ying, and K. Liu, "A study on predicting loan default based on the random forest algorithm," *Procedia Comput. Sci.*, vol. 162, pp. 503–513, 2019, doi: 10.1016/j.procs.2019.12.017.
- [5] D. Lin, Y. Li, and R. Zhang, "A SHAP-Based Interpretability Analysis of XGBoost for Loan Risk Prediction," in *Proceedings of the 2025 International Conference on Information Economy, Data Modeling and Cloud Computing*, New York, NY, USA: ACM, Aug. 2025, pp. 219–228. doi: 10.1145/3772900.3772936.
- [6] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3446–3453, Feb. 2012, doi: 10.1016/j.eswa.2011.09.033.
- [7] F. Weng, M. Zhu, M. Buckle, P. Hajek, and M. Z. Abedin, "Class imbalance Bayesian model averaging for consumer loan default prediction: The role of soft credit information," *Res. Int. Bus. Financ.*, vol. 74, p. 102722, Feb. 2025, doi: 10.1016/j.ribaf.2024.102722.
- [8] S. Shi, R. Tse, W. Luo, S. D'Addona, and G. Pau, "Machine learning-driven credit risk: a systemic review," *Neural Comput. Appl.*, vol. 34, no. 17, pp. 14327–14339, Sep. 2022, doi: 10.1007/s00521-022-07472-2.
- [9] F. O. Aghware *et al.*, "Enhancing the Random Forest Model via Synthetic Minority Oversampling Technique for Credit-Card Fraud Detection," *J. Comput. Theor. Appl.*, vol. 1, no. 4, pp. 407–420, Mar. 2024, doi: 10.62411/jcta.10323.
- [10] K. Wang, J. Wan, G. Li, and H. Sun, "A Hybrid Algorithm-Level Ensemble Model for Imbalanced Credit Default Prediction in the Energy Industry," *Energies*, vol. 15, no. 14, p. 5206, Jul. 2022, doi: 10.3390/en15145206.
- [11] M. Asutay and J. Othman, "Alternative measures for predicting financial distress in the case of Malaysian Islamic banks: assessing the impact of global financial crisis," *J. Islam. Account. Bus. Res.*, vol. 11, no. 10, pp. 1827–1845, Dec. 2020, doi: 10.1108/JIABR-12-2019-0223.
- [12] Y. Chen, R. Calabrese, and B. Martin-Barragan, "Interpretable machine learning for imbalanced credit scoring datasets," *Eur. J. Oper. Res.*, vol. 312, no. 1, pp. 357–372, Jan. 2024, doi: 10.1016/j.ejor.2023.06.036.
- [13] L. Yu, R. Zhou, R. Chen, and K. K. Lai, "Missing Data Preprocessing in Credit Classification: One-Hot Encoding or Imputation?," *Emerg. Mark. Financ. Trade*, vol. 58, no. 2, pp. 472–482, Jan. 2022, doi: 10.1080/1540496X.2020.1825935.
- [14] N. Alamsyah, B. Budiman, T. Parama Yoga, and R. Y. Rakhman Alamsyah, "A stacking ensemble

- model with SMOTE for improved imbalanced classification on credit data,” *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 22, no. 3, p. 657, Feb. 2024, doi: 10.12928/telkomnika.v22i3.25921.
- [15] B. Bischl *et al.*, “Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges,” *WIREs Data Min. Knowl. Discov.*, vol. 13, no. 2, Mar. 2023, doi: 10.1002/widm.1484.
 - [16] Y. Song and Y. Peng, “A MCDM-Based Evaluation Approach for Imbalanced Classification Methods in Financial Risk Prediction,” *IEEE Access*, vol. 7, pp. 84897–84906, 2019, doi: 10.1109/ACCESS.2019.2924923.
 - [17] T. Chen and C. Guestrin, “XGBoost,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
 - [18] M. Moscatelli, F. Parlapiano, S. Narizzano, and G. Viggiano, “Corporate default forecasting with machine learning,” *Expert Syst. Appl.*, vol. 161, p. 113567, Dec. 2020, doi: 10.1016/j.eswa.2020.113567.
 - [19] T. Shoko, T. Verster, and L. Dube, “Comparative analysis of classical and Bayesian optimisation techniques: Impact on model performance and interpretability in credit risk modelling using SHAP and PDPs,” *Data Sci. Financ. Econ.*, vol. 5, no. 3, pp. 320–354, 2025, doi: 10.3934/DSFE.2025014.
 - [20] M. Drehmann and M. Juselius, “Evaluating early warning indicators of banking crises: Satisfying policy requirements,” *Int. J. Forecast.*, vol. 30, no. 3, pp. 759–780, Jul. 2014, doi: 10.1016/j.ijforecast.2013.10.002.
 - [21] T. Saito and M. Rehmsmeier, “The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets,” *PLoS One*, vol. 10, no. 3, p. e0118432, Mar. 2015, doi: 10.1371/journal.pone.0118432.
 - [22] X. Zhu, Q. Chu, X. Song, P. Hu, and L. Peng, “Explainable prediction of loan default based on machine learning models,” *Data Sci. Manag.*, vol. 6, no. 3, pp. 123–133, Sep. 2023, doi: 10.1016/j.dsm.2023.04.003.
 - [23] H. Li and W. Wu, “Loan default predictability with explainable machine learning,” *Financ. Res. Lett.*, vol. 60, p. 104867, Feb. 2024, doi: 10.1016/j.frl.2023.104867.
 - [24] J. T. Hancock and T. M. Khoshgoftaar, “CatBoost for big data: an interdisciplinary review,” *J. Big Data*, vol. 7, no. 1, p. 94, Dec. 2020, doi: 10.1186/s40537-020-00369-8.