

OPTIMASI PENILAIAN KELAYAKAN PEMBIAYAAN SYARIAH MENGUNAKAN ALGORITMA MACHINE LEARNING (STUDI SIMULASI DENGAN DATA SINTETIS PROFIL ANGGOTA KOPERASI SYARIAH)

Dwi Wibowo¹, Ifan Ali Muntaha², Putut Ade Irawan^{3*}

^{1,2} Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pomosda Nganjuk Jawa Timur

³ Program Studi Teknik Industri Sekolah Tinggi Teknologi Pomosda Nganjuk Jawa Timur
wibowo.dwb@gmail.com

ABSTRAK

Pembiayaan bermasalah (*Non-Performing Financing/NPF*) masih menjadi tantangan bagi koperasi syariah, sementara penilaian kelayakan yang mengandalkan analisis 5C dan skoring manual rentan terhadap inkonsistensi. Penelitian ini membandingkan *Logistic Regression* (LR), *Random Forest* (RF), dan XGBoost untuk memprediksi keputusan kelayakan pembiayaan syariah melalui studi simulasi. Karena data anggota riil belum tersedia, digunakan 2.000 data pengajuan sintetis yang dibangkitkan dengan distribusi dan aturan keputusan yang dilaporkan secara lengkap, meliputi usia, pendapatan, jumlah tanggungan, plafon, tenor, jenis agunan, dan skor karakter. Label Layak/Tidak Layak menirukan keputusan kelayakan, bukan kejadian gagal bayar. Mengikuti kerangka CRISP-DM, data dibagi secara *stratified* 80:20 sebelum pra-pemrosesan; *One-Hot Encoding* dan standarisasi dijalankan dalam *pipeline* yang hanya di-fit pada data latih, dan *hyperparameter* dipilih dengan GridSearchCV 5-fold. Dengan kelas positif Layak, pada data uji LR memperoleh akurasi 82,0% dan AUC 0,898, RF *tuned* memperoleh akurasi 82,5% dan AUC 0,901, sedangkan XGBoost *tuned* memperoleh akurasi 83,25%, presisi 81,6%, *recall* 75,0%, F1-score 78,2%, dan AUC 0,910. Pada validasi silang, AUC ketiga model berada pada rentang 0,883–0,891 dengan selisih yang lebih kecil daripada simpangan bakunya. XGBoost menghasilkan performa terbaik pada data pengujian penelitian ini, tetapi keunggulannya tipis. Skor Karakter, jenis agunan, dan pendapatan bulanan merupakan fitur dengan kontribusi terbesar, sesuai rancangan simulasi. Hasil ini perlu dikonfirmasi dengan data koperasi riil sebelum model digunakan sebagai alat bantu keputusan.

Kata Kunci: Koperasi Syariah; Kelayakan Pembiayaan Syariah; Machine Learning; XGBoost; Feature Importance; Non-Performing Financing.

1. PENDAHULUAN

Lembaga keuangan syariah di Indonesia, khususnya di Provinsi Jawa Timur, menunjukkan pertumbuhan yang signifikan dalam satu dekade terakhir. Koperasi syariah hadir sebagai lembaga keuangan mikro yang vital bagi masyarakat menengah ke bawah yang belum sepenuhnya terlayani oleh perbankan formal. Namun, tingginya *Non-Performing Financing* (NPF) masih menjadi tantangan utama bagi lembaga keuangan syariah [1]. Pada koperasi syariah di Jawa Timur, rasio NPF dilaporkan berkisar antara 8–12%, di atas batas 5% yang lazim dijadikan acuan kesehatan pembiayaan oleh Otoritas Jasa Keuangan (OJK). Salah satu faktor penyebabnya adalah proses *underwriting* atau penilaian kelayakan calon anggota yang masih bersifat manual dan subjektif.

Penilaian kelayakan pembiayaan syariah secara konvensional bergantung pada prinsip 5C (*Character, Capacity, Capital, Collateral, Condition of Economy*) yang dilengkapi aspek kepatuhan syariah (5C+1S) [2]. Di koperasi syariah, proses ini sering kali berupa wawancara dan observasi sederhana oleh petugas lapangan, sehingga rentan terhadap bias dan inkonsistensi antarpetugas. Akibatnya, pengajuan dengan profil serupa dapat memperoleh keputusan yang berbeda.

Perkembangan *machine learning* dalam bidang keuangan menawarkan pendekatan yang lebih konsisten dan berbasis data untuk permasalahan ini [3].

Penelitian terdahulu menunjukkan keunggulan algoritma *ensemble* dalam klasifikasi kredit. Dumitrescu et al. [4] menunjukkan bahwa efek non-linear berbasis *decision tree* dapat meningkatkan kinerja regresi logistik dalam *credit scoring*, sedangkan Mienye dan Sun [5] merangkum keunggulan metode *ensemble* dalam menangani data yang kompleks dan mengandung *noise*. Pada lembaga keuangan mikro syariah, Mukit et al. [6] menunjukkan bahwa model berbasis pohon, termasuk *gradient boosting*, berpotensi meningkatkan penilaian kelayakan dengan memanfaatkan data non-finansial. Khusus untuk XGBoost, Gu et al. [7] menunjukkan bahwa model kartu skor berbasis XGBoost meningkatkan akurasi penilaian risiko kredit usaha mikro dan kecil, sedangkan Odunlami dan Nwonu [8] melaporkan bahwa metode *ensemble*, termasuk XGBoost dan LightGBM, yang dikombinasikan dengan teknik *resampling* memisahkan risiko gagal bayar lebih baik daripada model linear. Namun, kedua studi tersebut menggunakan label gagal bayar pada data non-koperasi. Kebaruan penelitian ini terletak pada

penerapan dan perbandingan XGBoost pada data koperasi syariah yang menggabungkan skor karakter terstruktur dengan data keuangan. Penelitian ini mengisi celah tersebut melalui studi simulasi yang dirancang menyerupai profil pengajuan pembiayaan pada koperasi syariah yang mayoritas anggotanya pelaku UMKM mikro, dengan dua batasan yang dinyatakan sejak awal: (1) data yang digunakan adalah data sintesis, bukan data anggota nyata, sehingga hasilnya menggambarkan kinerja model pada kondisi yang dirancang dan belum mencerminkan kinerja pada koperasi tertentu; dan (2) variabel target menirukan keputusan kelayakan, bukan status gagal bayar, sehingga model ini tidak dapat disamakan dengan model prediksi gagal bayar.

2. TINJAUAN PUSTAKA

2.1. Penilaian Pembiayaan Syariah

Pembiayaan syariah didasarkan pada akad yang melarang riba sesuai Al-Qur'an surat Al-Baqarah ayat 275–279 serta mengedepankan keadilan dan kejelasan (terhindar dari *gharar*). Prinsip penilaian kelayakan dalam pembiayaan syariah dikenal dengan 5C+1S, yang mencakup *Character* (karakter), *Capacity* (kapasitas), *Capital* (modal), *Collateral* (agunan), *Condition* (kondisi), dan *Syariah compliance* [2]. Dalam praktiknya, aspek karakter dan kapasitas umumnya menjadi pertimbangan utama petugas, sementara keterbatasan penilaian manual mendorong penggunaan *machine learning* sebagai alat bantu [3].

Secara konseptual, kapasitas membayar anggota dapat didekati dengan pendapatan per kapita rumah tangga sebagaimana Persamaan (1).

$$\text{Kapasitas} = \frac{\text{Pendapatan Bulanan}}{\text{Jumlah Tanggungan} + 1} \quad (1)$$

Secara teoretis, nilai kapasitas yang rendah menunjukkan ruang pembayaran angsuran yang lebih sempit. Proporsi pembiayaan yang diajukan terhadap pendapatan juga relevan: semakin besar proporsinya, semakin berat beban angsuran. Persamaan (1) digunakan sebagai kerangka konseptual dan tidak dimasukkan sebagai fitur tersendiri dalam model.

2.2. Machine Learning untuk Prediksi Kelayakan Kredit

Machine learning (ML) adalah cabang kecerdasan buatan yang memungkinkan sistem mempelajari pola dari data tanpa diprogram secara eksplisit. Dalam konteks keuangan, ML telah banyak digunakan untuk *credit scoring* dan evaluasi risiko kredit [3], termasuk untuk mitigasi NPF pada perbankan syariah [9].

2.3. Logistic Regression (LR)

LR merupakan algoritma klasifikasi biner yang memodelkan probabilitas kelas positif menggunakan fungsi sigmoid. Dalam penelitian ini,

$y = 1$ menyatakan kelas Layak, sehingga probabilitas suatu pengajuan diputuskan Layak untuk vektor fitur $\mathbf{x} = (x_1, \dots, x_p)$ dirumuskan pada Persamaan (2).

$$P(y = 1 | \mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \sum_{j=1}^p \beta_j x_j)}} \quad (2)$$

dengan β_0 adalah intersep, β_j adalah koefisien fitur ke- j , dan p adalah jumlah fitur. Pengajuan diklasifikasikan Layak apabila $P(y = 1 | \mathbf{x}) \geq 0,5$. Keunggulan LR terletak pada interpretabilitasnya yang tinggi, namun LR terbatas dalam menangkap hubungan non-linear dan interaksi antarfitur. Dumitrescu et al. [4] menunjukkan bahwa LR standar umumnya berkinerja lebih rendah daripada metode berbasis pohon dalam *credit scoring*, meskipun tetap menjadi acuan industri karena mudah dijelaskan.

2.4. Random Forest (RF)

RF adalah algoritma ensemble learning yang membangun banyak decision tree pada saat pelatihan dan menghasilkan prediksi berdasarkan majority voting [5]. RF efektif dalam menangani overfitting karena menggunakan teknik bagging (bootstrap aggregating) dan pemilihan fitur acak. Keunggulan lainnya adalah kemampuannya mengukur kepentingan fitur (feature importance) secara intrinsik. Mienye dan Sun [5] dalam IEEE Access mengkonfirmasi bahwa RF unggul dalam menangani dataset berdimensi tinggi dan mengandung noise, menjadikannya pilihan yang tepat untuk permasalahan *credit scoring*.

2.5. XGBoost (XGB)

Extreme Gradient Boosting (XGBoost) merupakan implementasi *gradient boosting* yang dioptimasi untuk kecepatan dan performa [10]. XGBoost membangun pohon secara berurutan; setiap pohon baru f_t mempelajari kesalahan model sebelumnya sehingga prediksi untuk data ke- i pada iterasi ke- t diperbarui menurut Persamaan (3).

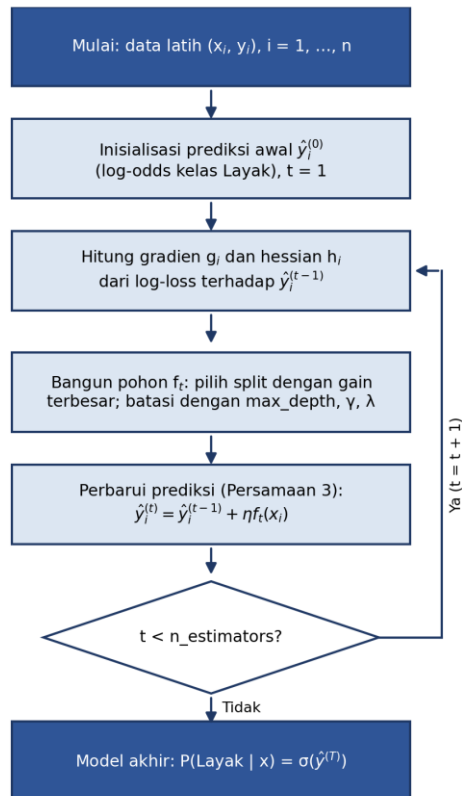
$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \quad (3)$$

dengan η adalah *learning rate*. Pohon f_t dipilih dengan meminimalkan fungsi objektif teregularisasi pada Persamaan (4) dan (5) [10].

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (4)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (5)$$

dengan $l(\cdot)$ adalah fungsi *loss* (pada penelitian ini *log-loss*), n adalah jumlah data latih, T adalah jumlah daun pohon, w_j adalah bobot daun ke- j , serta γ dan λ adalah parameter regularisasi yang mengendalikan kompleksitas pohon untuk mencegah *overfitting*. Mekanisme *boosting* yang iteratif dan regularisasi tersebut membuat XGBoost banyak digunakan pada data terstruktur, termasuk untuk penilaian risiko kredit usaha mikro dan kecil [7]. Alur pelatihan XGBoost dalam penelitian ini diringkas pada Gambar 1.



Gambar 1. Alur Pelatihan XGBoost

2.6. Studi Empiris Terkait

Beberapa penelitian terdahulu yang relevan dirangkum pada Tabel 1, gap yang diidentifikasi: mayoritas penelitian sebelumnya menggunakan data perbankan umum, *fintech*, atau lembaga keuangan mikro di luar Indonesia, dan belum mengkaji penggabungan skor karakter dengan data keuangan dalam konteks koperasi syariah. Penelitian ini mengkaji hal tersebut melalui studi simulasi sebagai tahap awal sebelum pengujian dengan data koperasi riil.

Tabel 1. Perbandingan Penelitian Terdahulu

Peneliti (Tahun)	Algoritma	Data	Hasil Utama	Gap
Odunlami & Nwonu [8] (2025)	LR, RF, XGB, LightGBM	Dataset kredit publik	Ensemble + resampling lebih baik dari model linear	Data bukan koperasi syariah; label default; tanpa skor karakter
Mukit et al. [6] (2026)	LinReg, RF, GB, Blockchain	Transaksi 1.275 petani (LKM syariah)	RF terbaik (regresi); GB: F1 0,91 (klasifikasi)	Data luar negeri; tanpa XGBoost dan skor karakter
Rawat [11] (2023)	CRISP-DM, RF, XGB	Data kredit bank	CRISP-DM efektif untuk pengembangan model kredit	Belum diuji pada koperasi syariah Indonesia
Mueankoo et al. [12] (2022)	LR, DT, RF, NB	Data pinjaman lembaga keuangan	RF: AUC 0,91; unggul atas LR	Konteks Thailand; tanpa skor karakter
Penelitian ini (2026)	LR, RF, XGB	2.000 data pengajuan sintesis (simulasi profil koperasi syariah)	XGBoost tuned: AUC 0,910 (tertinggi); selisih antarmodel kecil; target = keputusan kelayakan	—

3. METODE PENELITIAN

Metodologi penelitian mengacu pada *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang terdiri atas enam fase [11]: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan *deployment*. Pada penelitian ini fase *deployment* belum dilakukan; model hanya dievaluasi secara *offline* pada data historis.

3.1. Rancangan dan Pembangkitan Data Sintetis

Data anggota koperasi riil belum tersedia bagi peneliti. Oleh karena itu, penelitian ini menggunakan 2.000 data pengajuan pembiayaan

sintetis yang dibangkitkan dengan Python (NumPy, *seed* = 2026). Data sintetis tidak mewakili koperasi tertentu dan tidak memuat data pribadi. Rentang setiap variabel dipilih agar menyerupai profil anggota koperasi syariah skala mikro. Distribusi yang digunakan adalah sebagai berikut: usia berdistribusi seragam bilangan bulat 21–59 tahun; pendapatan bulanan berdistribusi lognormal dengan median Rp8,5 juta (simpangan baku logaritmik 0,35) dan dipotong pada Rp3–15 juta; jumlah tanggungan 0–4 orang dengan peluang 0,15; 0,25; 0,30; 0,20; dan 0,10; tenor 12, 24, 36, atau 48 bulan dengan peluang sama; plafon sama dengan pendapatan dikali bilangan acak seragam 2,5–9

ditambah galat normal (simpangan baku Rp5 juta), dipotong pada Rp5–100 juta; Skor Karakter bernilai $1 + 98 \times \text{Beta}(2, 2)$ yang dibulatkan; serta jenis agunan Sertifikat, BPKB, Emas, dan Tanpa Agunan dengan peluang 0,25; 0,35; 0,20; dan 0,20.

Status kelayakan ditentukan oleh skor laten z_i pada Persamaan (6). Pengajuan ke- i berstatus Layak apabila $z_i > 0$ dan Tidak Layak apabila sebaliknya. Deskripsi variabel disajikan pada Tabel 2.

$$z_i = \beta_0 + \beta_1(\text{SK}_i - 50) + \beta_2(P_i - 8,5) + \beta_3T_i + \beta_4\mathbb{1}(\text{DSR}_i > 0,40) + \beta_5\mathbb{1}(\text{SK}_i < 30) + \beta_6(U_i - 40) + \alpha_{A_i} + s\varepsilon_i \quad (6)$$

dengan SK_i adalah Skor Karakter, P_i adalah pendapatan (juta rupiah), T_i adalah jumlah

tanggungan, DSR_i adalah rasio angsuran terhadap pendapatan, yaitu (plafon/tenor)/pendapatan, U_i adalah usia, α_A adalah efek jenis agunan, dan ε_i adalah galat berdistribusi logistik standar. Nilai koefisien yang digunakan adalah $\beta_0 = 0,30$; $\beta_1 = 0,06$; $\beta_2 = 0,35$; $\beta_3 = -0,40$; $\beta_4 = -2,50$; $\beta_5 = -2,00$; $\beta_6 = 0,01$; $\alpha = 0,60$ (Sertifikat), 0,30 (BPKB), 0,40 (Emas), dan $-0,60$ (Tanpa Agunan); serta $s = 0,80$. Intersep β_0 dipilih agar proporsi kelas Layak sekitar 40%. Aturan ini merupakan asumsi rancangan yang merepresentasikan aspek karakter, kapasitas, dan agunan dalam prinsip 5C, termasuk efek ambang yang tidak linear ($\text{DSR} > 0,40$ dan Skor Karakter < 30). Aturan ini bukan temuan empiris. Kode pembangkit data tersedia dari penulis.

Tabel 2. Deskripsi Variabel Dataset

Variabel	Tipe Data	Deskripsi	Contoh Nilai
Usia	Numerik (Integer)	Usia anggota (tahun)	21–59
Pendapatan_Bulanan	Numerik (Integer)	Pendapatan bersih per bulan (IDR)	Rp3.000.000–Rp15.000.000
Jumlah_Tanggungan	Numerik (Integer)	Jumlah anggota keluarga yang ditanggung	0–4
Jumlah_Pembiayaan	Numerik (Integer)	Plafon pembiayaan yang diajukan (IDR)	Rp5.000.000–Rp100.000.000
Durasi_Bulan	Numerik (Integer)	Tenor pembiayaan (bulan)	12, 24, 36, 48
Skor_Karakter	Numerik (Integer)	Skor karakter simulasi, skala 1–99 (Subbab 3.3)	1–99
Agunan_Tersedia	Kategorik (String)	Jenis jaminan (diolah dengan One-Hot Encoding)	BPKB, Emas, Sertifikat, Tanpa Agunan
Status_Kelayakan	Target (Binary)	Keputusan kelayakan hasil simulasi (Pers. 6), bukan status pembayaran	Layak (1); Tidak Layak (0)

3.2. Definisi Variabel Target dan Kelas Positif

Variabel target Status_Kelayakan menirukan keputusan kelayakan atas setiap pengajuan: Layak (1) berarti pengajuan layak disetujui, dan Tidak Layak (0) berarti pengajuan layak ditolak. Label ini dibangkitkan dari aturan pada Persamaan (6) dan tidak berasal dari riwayat pembayaran. Dengan demikian, model dalam penelitian ini memprediksi keputusan kelayakan dan tidak memprediksi gagal bayar (*default*). Istilah “risiko tinggi” dan “gagal bayar” tidak digunakan sebagai nama kelas; kedua istilah tersebut hanya dipakai saat membahas NPF atau penelitian lain yang memang menggunakan label pembayaran.

Kelas positif pada seluruh metrik adalah Layak (1). *True positive* (TP) adalah pengajuan Layak yang diprediksi Layak; *true negative* (TN) adalah pengajuan Tidak Layak yang diprediksi Tidak Layak; *false positive* (FP) adalah pengajuan Tidak Layak yang diprediksi Layak, yaitu kesalahan yang berpotensi meloloskan pengajuan yang semestinya ditolak; dan *false negative* (FN) adalah pengajuan Layak yang diprediksi Tidak Layak, yaitu kesalahan yang berpotensi menolak anggota yang semestinya disetujui. Presisi, *recall*, dan *F1-score* dihitung untuk kelas Layak, sedangkan AUC dihitung dari probabilitas prediksi kelas Layak.

3.3. Variabel Skor Karakter

Dalam praktik koperasi, skor karakter umumnya diperoleh dari penilaian petugas terhadap calon anggota, misalnya melalui wawancara dan informasi dari lingkungan. Dalam simulasi ini, Skor Karakter merupakan variabel sintetis berskala 1–99 yang merepresentasikan hasil penilaian tersebut, tanpa instrumen atau penilai nyata. Karena Skor Karakter dirancang memiliki pengaruh terbesar dalam aturan keputusan (Persamaan 6), dominansinya dalam model merupakan konsekuensi rancangan. Penerapan pada data riil memerlukan instrumen yang tervalidasi, dengan jumlah butir, skala, penilai, dan reliabilitas yang jelas, karena penilaian petugas berpotensi subjektif.

3.4. Pra-Pemrosesan dan Pencegahan Data Leakage

Pra-pemrosesan dilakukan dengan Python menggunakan pustaka *scikit-learn* [13] dan *XGBoost* [10]. Untuk mencegah *data leakage* [14], urutan kerjanya adalah sebagai berikut.

a. Pembagian data.

Pembagian data dilakukan paling awal. Data dibagi secara *stratified* menjadi 80% data latih (1.600 pengajuan: 642 Layak dan 958 Tidak Layak) dan 20% data uji (400 pengajuan: 160 Layak dan 240 Tidak Layak) dengan *random_state* = 42, sehingga proporsi kelas Layak dan Tidak Layak pada kedua subset sama

dengan data asal. Data uji tidak digunakan dalam pra-pemrosesan, pemilihan *hyperparameter*, maupun pemilihan model.

b. Encoding data kategorikal.

Fitur Agunan_Tersedia diubah menggunakan *One-Hot Encoding* menjadi empat variabel biner (Sertifikat, BPKB, Emas, Tanpa Agunan). *Ordinal Encoding* tidak digunakan karena urutan antarkategori tidak memiliki dasar yang kuat: sertifikat tanah/bangunan umumnya bernilai paling tinggi, tetapi emas justru paling likuid, sehingga urutan Sertifikat > BPKB > Emas > Tanpa Agunan akan memaksakan urutan dan jarak yang sama antarkategori. *One-Hot Encoding* tidak mengasumsikan urutan maupun jarak antarkategori.

c. Standarisasi.

Untuk LR yang sensitif terhadap skala, fitur numerik distandarisasi dengan *StandardScaler* menurut Persamaan (7).

$$z = \frac{x - \mu_{\text{latih}}}{\sigma_{\text{latih}}} \quad (7)$$

dengan μ_{latih} dan σ_{latih} adalah rata-rata dan simpangan baku yang dihitung hanya dari data latih, lalu diterapkan tanpa perubahan pada data uji. RF dan XGBoost tidak memerlukan standarisasi karena pemisahan pada pohon keputusan tidak bergantung pada skala fitur.

d. Pipeline.

Encoder, *scaler*, dan model dibungkus dalam satu objek *Pipeline*. Dengan cara ini, saat *GridSearchCV* dan validasi silang dijalankan, encoder dan *scaler* di-fit ulang hanya pada *fold* latih di setiap iterasi dan tidak pernah melihat *fold* validasi maupun data uji.

3.5. Perancangan Model

Tiga model *baseline* dibangun dengan konfigurasi awal berikut: (1) LR: solver = 'liblinear', max_iter = 1000; (2) RF: n_estimators = 100, max_depth = None, random_state = 42; dan (3) XGBoost: n_estimators = 100, learning_rate = 0,3, max_depth = 6, eval_metric = 'logloss', random_state = 42. Ketiga model dilatih pada data latih yang sama melalui *pipeline* pada Subbab 3.4.

3.6. Hyperparameter Tuning dan Validasi Silang

Hyperparameter dipilih menggunakan *GridSearchCV* dengan *5-fold stratified cross-validation* (shuffle = True, random_state = 42) pada data latih dengan metrik pemilihan ROC-AUC. Ruang pencarian untuk LR adalah $C \in \{0,01; 0,1; 1; 10; 100\}$; untuk RF adalah $n_{\text{estimators}} \in \{50, 100, 200\}$ dan $\text{max_depth} \in \{10, 20, \text{None}\}$; dan untuk XGBoost adalah $\text{learning_rate} \in \{0,01; 0,1; 0,3\}$ dan $\text{max_depth} \in \{3, 6, 9\}$ dengan $n_{\text{estimators}} = 100$. Kombinasi terbaik di-fit ulang pada seluruh data latih dan dievaluasi satu kali pada data uji.

Untuk menilai stabilitas, konfigurasi terbaik tiap model juga divalidasi silang *5-fold* pada data latih, dan hasilnya dilaporkan sebagai rata-rata $\pm$ simpangan baku untuk akurasi, presisi, *recall*, F1-score, dan AUC. Model akhir dipilih berdasarkan AUC validasi silang tertinggi.

3.7. Metrik Evaluasi

Evaluasi dilakukan dalam tiga tahap: (1) menyusun *confusion matrix* setiap model pada data uji; (2) menghitung metrik turunan dari *confusion matrix* dan probabilitas prediksi; serta (3) membandingkan hasil data uji dengan hasil validasi silang untuk menilai stabilitas. Dengan kelas positif Layak (Subbab 3.2), metrik dihitung menurut Persamaan (8)–(11).

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (10)$$

$$F1 = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (11)$$

AUC (*area under the ROC curve*) mengukur kemampuan model memberikan probabilitas kelas Layak yang lebih tinggi kepada pengajuan Layak dibandingkan pengajuan Tidak Layak. AUC merupakan luas di bawah kurva ROC (Persamaan 12) dan secara ekuivalen dapat dihitung dengan Persamaan (13).

$$AUC = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR}) \quad (12)$$

$$AUC = \frac{1}{n_1 n_0} \sum_{i \in P} \sum_{j \in N} \mathbb{1}(s_i > s_j) \quad (13)$$

dengan $\text{TPR} = \text{TP}/(\text{TP} + \text{FN})$, $\text{FPR} = \text{FP}/(\text{FP} + \text{TN})$, P dan N adalah himpunan pengajuan Layak dan Tidak Layak pada data uji (berukuran n_1 dan n_0), s_i adalah probabilitas kelas Layak untuk pengajuan ke- i , dan $\mathbb{1}(\cdot)$ bernilai 1 jika kondisinya benar dan 0 jika tidak (pasangan dengan skor sama diberi bobot 0,5).

3.8. Etika Penelitian dan Ketersediaan Data

Penelitian ini tidak menggunakan data anggota koperasi nyata. Seluruh data dibangkitkan secara sintetis dan tidak ada nilai yang disalin atau diturunkan dari catatan individu, sehingga data tidak memuat informasi yang dapat mengidentifikasi seseorang dan tidak termasuk pemrosesan data pribadi sebagaimana diatur dalam Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi [15]. Penerapan model pada data koperasi riil di kemudian hari memerlukan izin tertulis pengelola data, anonimisasi, pembatasan akses, dan pelaporan hasil dalam bentuk agregat. Karena keputusan pembiayaan berdampak langsung pada anggota, model semacam ini hanya layak diposisikan sebagai alat bantu, dengan keputusan akhir tetap berada pada Komite pembiayaan. Data sintetis dan kode analisis tersedia dari penulis atas permintaan.

4. HASIL DAN PEMBAHASAN

4.1. Analisis Eksplorasi Data (EDA)

Dari 2.000 data pengajuan sintetis, sebanyak 802 (40,1%) berstatus Layak dan 1.198 (59,9%) berstatus Tidak Layak. Rasio kelas sekitar 40:60

menunjukkan ketidakseimbangan moderat; karena itu, selain akurasi, penelitian ini melaporkan presisi, *recall*, *F1-score*, *AUC*, dan *confusion matrix*, serta menggunakan *stratified split*. Statistik deskriptif variabel numerik disajikan pada Tabel 3.

Tabel 3. Statistik Deskriptif Variabel Numerik

Variabel	Rata-rata	Std. Dev	Min	Max
Usia (tahun)	40,3	11,2	21	59
Pendapatan_Bulanan (juta IDR)	8,75	2,88	3,00	15,00
Jumlah_Tanggungan	1,9	1,2	0	4
Jumlah_Pembiayaan (juta IDR)	49,4	23,0	5,0	100,0
Durasi_Bulan	30,6	13,5	12	48
Skor_Karakter (1–99)	50,5	22,1	2	98

Pengajuan berstatus Layak memiliki rata-rata pendapatan (Rp9,91 juta) dan skor karakter (63,9 poin) yang lebih tinggi dibandingkan pengajuan Tidak Layak (Rp7,98 juta; 41,5 poin). Proporsi Layak pada pengajuan tanpa agunan (30,9%) lebih rendah daripada pengajuan beragunan (41,7–42,9%). Sebanyak 15,6% pengajuan memiliki DSR > 0,40 dan 20,5% memiliki Skor Karakter < 30. Korelasi antara Jumlah_Pembiayaan dan

Pendapatan_Bulanan bernilai positif (0,67), sesuai rancangan pembangkitan plafon. EDA hanya bersifat deskriptif dan tidak digunakan untuk menentukan parameter pra-pemrosesan.

4.2. Performa Model pada Data Uji

Perbandingan performa model *baseline* dan *tuned* pada data uji (400 pengajuan; kelas positif = Layak) disajikan pada Tabel 4.

Tabel 4. Perbandingan Metrik Evaluasi Model pada Data Uji

Model	Akurasi	Presisi	Recall	F1-Score	ROC-AUC
Logistic Regression (Baseline)	0,820	0,814	0,713	0,760	0,898
Logistic Regression (Tuned)	0,820	0,814	0,713	0,760	0,898
Random Forest (Baseline)	0,835	0,813	0,762	0,787	0,900
Random Forest (Tuned)	0,825	0,817	0,725	0,768	0,901
XGBoost (Baseline)	0,810	0,780	0,731	0,755	0,882
XGBoost (Tuned)*	0,833	0,816	0,750	0,782	0,910

* Model terpilih berdasarkan AUC validasi silang tertinggi (Tabel 7). Akurasi XGBoost (Tuned) = $333/400 = 0,8325$.

Pada Tabel 4, ketiga model menghasilkan performa yang berdekatan. LR memperoleh akurasi 82,0% dan AUC 0,898; *tuning* tidak mengubah hasilnya karena nilai C terbaik (100) menghasilkan model yang praktis sama dengan konfigurasi awal. RF *tuned* memperoleh akurasi 82,5% dan AUC 0,901, sedangkan XGBoost *tuned* memperoleh akurasi 83,25%, presisi 81,6%, *recall* 75,0%, *F1-score* 78,2%, dan AUC 0,910. Di antara model *tuned*, XGBoost menghasilkan performa terbaik pada data pengujian penelitian ini untuk akurasi, *recall*, *F1-score*, dan AUC, tetapi selisih akurasinya hanya 1,25 poin persentase terhadap LR (*baseline* maupun *tuned*) dan 0,75 poin persentase terhadap RF *tuned*. RF *baseline* bahkan memperoleh akurasi (83,5%) dan *F1-score* (78,7%) sedikit lebih tinggi daripada XGBoost *tuned*. Hal ini menunjukkan bahwa pemilihan *hyperparameter* berdasarkan AUC validasi silang tidak selalu meningkatkan setiap metrik pada satu kali pembagian data uji. Oleh karena itu, perbandingan model perlu dibaca bersama hasil validasi silang pada Subbab 4.4.

4.3. Confusion Matrix

Tabel 5 menyajikan *confusion matrix* ketiga model *tuned* pada data uji. Kelas aktual dan kelas prediksi mengikuti definisi pada Subbab 3.2 dengan kelas positif Layak. FP menunjukkan jumlah pengajuan Tidak Layak yang diprediksi Layak, sedangkan FN menunjukkan jumlah pengajuan Layak yang diprediksi Tidak Layak.

Tabel 5. Confusion Matrix Model Tuned pada Data Uji

Model	TN	FP	FN	TP	Total
Logistic Regression	214	26	46	114	400
Random Forest	214	26	44	116	400
XGBoost	213	27	40	120	400

TN = aktual Tidak Layak, prediksi Tidak Layak; FP = aktual Tidak Layak, prediksi Layak; FN = aktual Layak, prediksi Tidak Layak; TP = aktual Layak, prediksi Layak. Untuk setiap model, TN + FP = 240 (aktual Tidak Layak) dan FN + TP = 160 (aktual Layak).

Bagi koperasi, FP lebih berisiko karena dapat meloloskan pengajuan yang semestinya ditolak, sedangkan FN berpotensi mengurangi akses pembiayaan bagi anggota yang layak. Jumlah FP

ketiga model hampir sama (26–27), sedangkan XGBoost menghasilkan FN paling sedikit (40), dibandingkan RF (44) dan LR (46). Keunggulan XGBoost pada data uji terutama berasal dari berkurangnya FN, yaitu pengajuan Layak yang salah ditolak. Jika koperasi ingin menekan FP, ambang probabilitas klasifikasi dapat dinaikkan di atas 0,5 dengan konsekuensi bertambahnya FN.

4.4. Hasil Hyperparameter Tuning dan Validasi Silang

Tabel 6 menyajikan ruang pencarian dan parameter terbaik hasil GridSearchCV, sedangkan Tabel 7 menyajikan hasil validasi silang 5-fold pada data latih untuk konfigurasi terbaik setiap model.

Tabel 6. Ruang Pencarian dan Parameter Terbaik GridSearchCV

Model	Parameter	Nilai yang diuji	Nilai terbaik
Logistic Regression	C	0,01; 0,1; 1; 10; 100	100
Random Forest	n_estimators	50; 100; 200	200
	max_depth	10; 20; None	10
XGBoost	learning_rate	0,01; 0,1; 0,3	0,1
	max_depth	3; 6; 9	3

Tabel 7. Hasil Validasi Silang 5-Fold pada Data Latih (Rata-rata $\pm$ SD)

Metrik	LR	RF	XGBoost
Akurasi	0,806 $\pm$ 0,027	0,802 $\pm$ 0,014	0,810 $\pm$ 0,014
Presisi	0,770 $\pm$ 0,038	0,764 $\pm$ 0,028	0,767 $\pm$ 0,030
Recall	0,737 $\pm$ 0,043	0,735 $\pm$ 0,053	0,760 $\pm$ 0,050
F1-score	0,752 $\pm$ 0,035	0,748 $\pm$ 0,024	0,762 $\pm$ 0,022
AUC	0,886 $\pm$ 0,014	0,883 $\pm$ 0,007	0,891 $\pm$ 0,007

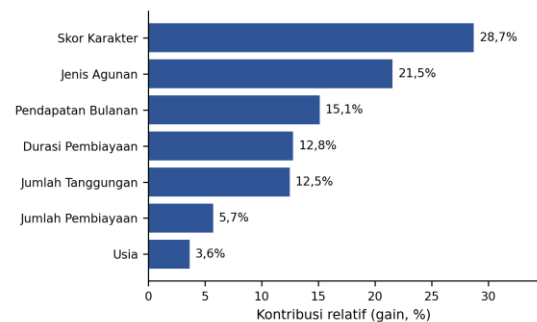
Setiap fold validasi berisi 320 pengajuan (128–129 Layak); SD dihitung dari lima fold.

Pada validasi silang, XGBoost memperoleh rata-rata AUC tertinggi (0,891 $\pm$ 0,007), diikuti LR (0,886 $\pm$ 0,014) dan RF (0,883 $\pm$ 0,007). Rata-rata akurasi ketiga model berada pada rentang 0,802–0,810. Selisih antarmodel lebih kecil daripada simpangan bakunya, sehingga keunggulan XGBoost pada data ini tergolong tipis dan belum dapat dinyatakan berbeda secara bermakna. Nilai validasi silang sedikit lebih rendah daripada hasil data uji, yang menunjukkan adanya variasi akibat pembagian data. Kinerja LR yang mendekati model berbasis pohon menunjukkan bahwa sebagian besar pola dalam data sintetis dapat ditangkap secara linear,

sedangkan efek ambang pada Persamaan (6) hanya memberi sedikit keuntungan bagi model non-linear.

4.5. Analisis Feature Importance

Gambar 2 menunjukkan *feature importance* model XGBoost *tuned* berdasarkan *gain*, yaitu rata-rata peningkatan fungsi objektif ketika suatu fitur digunakan sebagai pemisah. Nilai untuk jenis agunan merupakan penjumlahan *gain* dari keempat variabel hasil *One-Hot Encoding*.



Gambar 2. Feature Importance Model XGBoost (Tuned)

Urutan kontribusi fitur adalah Skor Karakter (28,7%), Jenis Agunan (21,5%), Pendapatan Bulanan (15,1%), Durasi Pembiayaan (12,8%), Jumlah Tanggungan (12,5%), Jumlah Pembiayaan (5,7%), dan Usia (3,6%). Urutan ini sebagian besar mencerminkan rancangan simulasi. Skor Karakter memiliki koefisien dan efek ambang terbesar pada Persamaan (6). Jenis agunan memperoleh *gain* tinggi karena kategori Tanpa Agunan memiliki proporsi Layak yang jauh lebih rendah. Durasi dan jumlah pembiayaan berpengaruh melalui DSR, sehingga kontribusinya terbagi antara kedua variabel tersebut. Usia memiliki kontribusi terkecil karena koefisiennya sangat kecil. Dengan demikian, analisis ini menunjukkan bahwa XGBoost mampu mengenali variabel yang memang dirancang berpengaruh, tetapi tidak dapat dijadikan bukti empiris tentang faktor penentu kelayakan di koperasi. *Feature importance* juga tidak menunjukkan hubungan sebab-akibat.

4.6. Pembahasan

AUC XGBoost sebesar 0,910 menunjukkan bahwa model mampu membedakan pengajuan Layak dan Tidak Layak dengan baik pada data uji sintetis. Batas atas kinerja dibatasi oleh galat acak dalam aturan keputusan (Persamaan 6), sehingga akurasi di sekitar 80–83% merupakan konsekuensi rancangan dan bukan kelemahan model semata. Karena label menirukan keputusan kelayakan, hasil ini belum membuktikan bahwa model mampu memprediksi gagal bayar atau menurunkan NPF.

Sebagai ilustrasi, calon anggota berusia 35 tahun dengan pendapatan Rp5.000.000, tiga tanggungan, plafon Rp60.000.000, tenor 12 bulan

(DSR = 1,0), Skor Karakter 20, dan agunan BPKB diklasifikasikan Tidak Layak oleh XGBoost dengan probabilitas kelas Tidak Layak 99,6%, dan oleh LR dengan probabilitas 99,8%. Kedua model menghasilkan keputusan yang sama karena profil ini memicu dua efek ambang sekaligus ($DSR > 0,40$ dan Skor Karakter < 30). Perbedaan antarmodel lebih terlihat pada profil yang berada di sekitar batas keputusan.

Keunggulan tipis XGBoost sejalan dengan Mienye dan Sun [5] mengenai metode *ensemble* dan dengan Mukit et al. [6] yang menunjukkan potensi *machine learning* untuk *credit scoring* di lembaga keuangan mikro syariah. Namun, hasil simulasi ini juga menunjukkan bahwa LR tetap kompetitif ketika pola data sebagian besar linear. Dibandingkan Rawat [11] yang menerapkan CRISP-DM pada data kredit bank, penelitian ini menyediakan kerangka evaluasi yang bebas *data leakage* dan dapat direplikasi, serta dapat langsung diterapkan ketika data koperasi riil tersedia.

5. KESIMPULAN DAN SARAN

Penelitian ini membandingkan LR, RF, dan XGBoost untuk memprediksi keputusan kelayakan pembiayaan syariah melalui studi simulasi dengan 2.000 data pengajuan sintetis. Kesimpulannya adalah: (1) XGBoost *tuned* menghasilkan performa terbaik pada data pengujian penelitian ini dengan akurasi 83,25%, presisi 81,6%, *recall* 75,0%, F1-score 78,2%, dan AUC 0,910, diikuti RF *tuned* (akurasi 82,5%; AUC 0,901) dan LR (akurasi 82,0%; AUC 0,898); (2) keunggulan XGBoost tipis, yaitu 1,25 poin persentase akurasi terhadap LR dan 0,75 poin persentase terhadap RF *tuned*, dan pada validasi silang selisih antarmodel lebih kecil daripada simpangan bakunya; dan (3) Skor Karakter, jenis agunan, dan pendapatan bulanan merupakan fitur dengan kontribusi terbesar, sesuai rancangan simulasi.

Penelitian ini memiliki beberapa keterbatasan. Pertama, data bersifat sintetis dan aturan keputusannya ditetapkan oleh peneliti, sehingga hasil kinerja dan *feature importance* mencerminkan rancangan simulasi dan tidak dapat digeneralisasi ke koperasi mana pun. Kedua, model belum diuji dengan data riil maupun validasi eksternal. Ketiga, Skor Karakter merupakan variabel simulasi; pada praktiknya skor ini dinilai petugas dan berpotensi subjektif. Keempat, label menirukan keputusan kelayakan, bukan status pembayaran, sehingga model tidak dapat dipakai untuk memprediksi gagal bayar. Oleh karena itu, model belum boleh digunakan sebagai dasar persetujuan pembiayaan, apalagi sebagai satu-satunya dasar.

Penelitian selanjutnya disarankan untuk (1) menerapkan kerangka ini pada data riil dari satu atau beberapa koperasi syariah, disertai izin dan anonimisasi, serta melakukan validasi eksternal dan validasi berbasis waktu (*out-of-time*); (2)

menggunakan label status pembayaran, misalnya tunggakan lebih dari 90 hari, agar model dapat diarahkan pada prediksi gagal bayar; (3) mengembangkan dan memvalidasi instrumen Skor Karakter, termasuk reliabilitas dan konsistensi antarpemilai; dan (4) menambahkan uji signifikansi perbedaan antarmodel, analisis interpretabilitas per kasus, dan evaluasi keadilan (*fairness*).

DAFTAR PUSTAKA

- [1] Fitri, S. D., & Sriyana, J. (2023). Analisis faktor-faktor yang memengaruhi tingkat non-performing financing (NPF) di Bank Umum Syariah Indonesia periode 2015–2021. *Jurnal Informatika Ekonomi Bisnis*, 5(1), 232–239. <https://doi.org/10.37034/infeb.v5i1.240>
- [2] Asy'ari, M. S., & Wigati, S. (2024). Analisis kelayakan nasabah menggunakan prinsip 5C 1S dalam pembiayaan pemilikan rumah (PPR) (Studi pada KB Bank Syariah KC Sidoarjo). *Jurnal Ilmiah Ekonomi dan Manajemen*, 2(11). <https://doi.org/10.61722/jiem.v2i11.2921>
- [3] Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: A systematic literature review. *Journal of Banking and Financial Technology*, 4(1), 111–138. <https://doi.org/10.1007/s42786-020-00020-3>
- [4] Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 297(3), 1178–1192. <https://doi.org/10.1016/j.ejor.2021.06.053>
- [5] Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, 10, 99129–99149. <https://doi.org/10.1109/ACCESS.2022.3207287>
- [6] Mukit, M. M. H., Hasan, F., Choudhury, T., Al Fadli, A., & Fadul, A. (2026). Machine learning & artificial intelligence powered credit scoring models for Islamic microfinance institutions: A blockchain approach. *Risks*, 14(1), 12. <https://doi.org/10.3390/risks14010012>
- [7] Gu, Z., Lv, J., Wu, B., Hu, Z., & Yu, X. (2024). Credit risk assessment of small and micro enterprise based on machine learning. *Heliyon*, 10(5), e27096. <https://doi.org/10.1016/j.heliyon.2024.e27096>
- [8] Odunlami, B. G., & Nwonu, B. (2025). Credit risk prediction using ensemble and linear machine learning models. *Journal of Advanced Artificial Intelligence*, 2(1), 1–8. <https://doi.org/10.5120/jaai202441>
- [9] Hasanah, E. N., Mudrikah, A., & Mahfuzh, M. A. (2025). The role of artificial intelligence in reducing non-performing financing (NPF) in Islamic banking: A systematic literature review.

- Proceeding International Seminar of Islamic Studies.*
<https://jurnal.umsu.ac.id/index.php/insis/article/view/23099>
- [10] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
- [11] Rawat, K. (2023). Applying CRISP-DM methodology in developing machine learning model for credit risk prediction. In K. Arai (Ed.), *Intelligent Computing: SAI 2023* (Lecture Notes in Networks and Systems, Vol. 739, pp. 522–538). Springer. https://doi.org/10.1007/978-3-031-37963-5_37
- [12] Mueankoo, A., Eso, M., & Musikasuwan, S. (2022). Performance of a loan repayment status model using machine learning. In *2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. IEEE. <https://doi.org/10.1109/ECTI-CON54298.2022.9795570>
- [13] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://jmlr.org/papers/v12/pedregosa11a.html>
- [14] Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data*, 6(4), Article 15. <https://doi.org/10.1145/2382577.2382579>
- [15] Republik Indonesia. (2022). *Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi*. Jakarta: Sekretariat Negara.