

**PREDIKSI RESIKO KREDIT CALON NASABAH PADA
LAYANAN PEER-TO-PEER LENDING BERBASIS MACHINE
LEARNING: KOMPARASI MODEL RANDOM FOREST DAN
NAIVE BAYES**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

AGNES KAMELIA NARJUN

22.12.2270

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**PREDIKSI RESIKO KREDIT CALON NASABAH PADA
LAYANAN PEER-TO-PEER LENDING BERBASIS MACHINE
LEARNING: KOMPARASI MODEL RANDOM FOREST DAN
NAIVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

AGNES KAMELIA NARJUN

22.12.2270

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**PREDIKSI RESIKO KREDIT CALON NASABAH PADA
LAYANAN PEER-TO-PEER LENDING BERBASIS MACHINE
LEARNING: KOMPARASI MODEL RANDOM FOREST DAN
NAIVE BAYES**

yang disusun dan diajukan oleh

Agnes Kamelia Narjun

22.12.2270

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 18 Februari 2026

Dosen Pembimbing,



Krisnawati, S.Si., M.T.
NIK. 190302038

HALAMAN PENGESAHAN

SKRIPSI

**PREDIKSI RESIKO KREDIT CALON NASABAH PADA
LAYANAN PEER-TO-PEER LENDING BERBASIS MACHINE
LEARNING: KOMPARASI MODEL RANDOM FOREST DAN
NAIVE BAYES**

yang disusun dan diajukan oleh

Agnes Kamelia Narjun

22.12.2270

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Ali Mustopa, S.Kom., M.Kom
NIK. 190302192

Norhikmah, S.Kom., M.Kom
NIK. 190302245

Krisnawati, S.Si., M.T
NIK. 190302106

Tanda Tangan

 5/2/26

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Agnes Kamelia Narjun
NIM : 22.12.2270

Menyatakan bahwa Skripsi dengan judul berikut:

Prediksi Resiko Kredit Calon Nasabah Pada Layanan Peer-To-Peer Lending Berbasis Machine Learning: Komparasi Model Random Forest dan Naïve Bayes

Dosen Pembimbing : Krisnawati, S.Si., M.T

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 18 Februari 2026

Yang Menyatakan,



Agnes Kamelia Narjun

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada program studi yang penulis tempuh.

Dalam proses penyusunan skripsi ini, penulis memperoleh banyak bimbingan, arahan, dukungan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Ibu Krisnawati, S.Si., M.T. selaku Dosen Pembimbing yang telah meluangkan waktu, memberikan bimbingan, arahan, serta motivasi dengan penuh kesabaran dan tanggung jawab selama proses penyusunan skripsi ini.
2. Tim Dosen Penguji yang telah memberikan saran, masukan, dan evaluasi yang konstruktif demi penyempurnaan skripsi ini.
3. Orang tua serta keluarga penulis yang senantiasa memberikan doa, dukungan moral, dan semangat sehingga penulis dapat menyelesaikan studi dengan baik.
4. Semua pihak yang telah membantu secara langsung maupun tidak langsung dalam proses penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun demi perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan pihak-pihak yang membutuhkan

Yogyakarta, 18 Februari 2026

Agnes Kamelia Narjun

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH.....	xv
INTISARI	xvii
ABSTRACT.....	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7

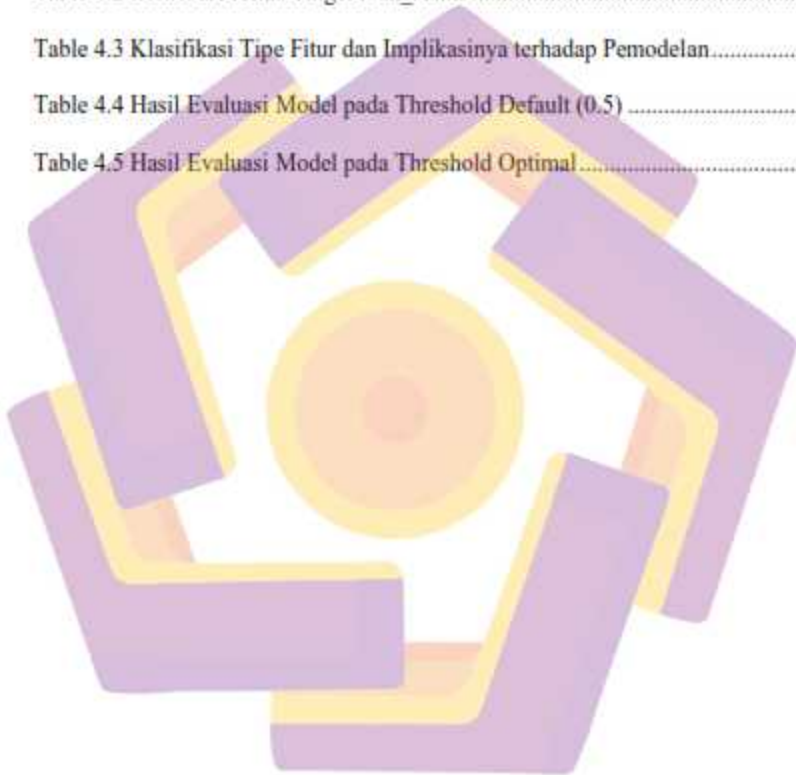
2.1	Studi Literatur	7
2.2	Dasar Teori.....	19
2.2.1	Konsep Dasar Risiko Kredit dan Default.....	19
2.2.2	Layanan Peer-to-Peer Lending dan Karakteristik Risikonya.....	21
2.2.3	Konsep Dasar Machine Learning.....	23
2.2.4	Pra-pemrosesan Data (<i>Data Preprocessing</i>).....	24
2.2.5	Algoritma Naive Bayes	27
2.2.6	Algoritma Random Forest.....	28
2.2.7	Prinsip Validasi Model Menggunakan Stratified Cross-Validation	29
2.2.8	Evaluasi Model Klasifikasi pada Data Tidak Seimbang.....	30
BAB III METODE PENELITIAN		34
3.1	Objek Penelitian.....	34
3.2	Alur Penelitian	34
3.2.1	Pengumpulan dan Pemahaman Data.....	35
3.2.2	Sampling Dataset	36
3.2.3	Seleksi Fitur	36
3.2.4	Pra-pemrosesan Data	39
3.2.5	Pembagian Tipe Fitur dan Perancangan Pipeline Model.....	41
3.2.6	Stratified 5-Fold Cross-Validation.....	43
3.2.7	Pelatihan Model	43
3.2.8	Evaluasi Model	44
3.2.9	Analisis Perbandingan Model.....	45

3.3	Alat dan Bahan.....	45
BAB IV HASIL DAN PEMBAHASAN		47
4.1	Pengumpulan, Pemahaman, dan Karakteristik Dataset	47
4.1.1	Sumber dan Proses Pembentukan Dataset	47
4.1.2	Karakteristik dan Klasifikasi Fitur Prediktor	47
4.1.3	Ringkasan Fitur yang Digunakan.....	48
4.1.4	Variabel target dan Distribusi Kelas	51
4.2	Hasil Pra-pemrosesan dan Pembagian Fitur	52
4.3	Konfigurasi Model dan Proses Pelatihan	54
4.3.1	Tujuan Konfigurasi dan Pelatihan Model	54
4.3.2	Konfigurasi Gaussian Naive Bayes	54
4.3.3	Konfigurasi Random Forest.....	55
4.3.4	Skema Validasi dan Proses Pelatihan	56
4.3.5	Implikasi Konfigurasi terhadap Komparasi Model	56
4.4	Hasil Evaluasi Model pada Threshold Default (0.5).....	57
4.4.1	Tujuan Evaluasi pada Threshold Default.....	57
4.4.2	Hasil Evaluasi Gaussian Naive Bayes	57
4.4.3	Hasil Evaluasi Random Forest	58
4.4.4	Analisis Perbandingan Antar Model	59
4.4.5	Implikasi terhadap Deteksi Risiko Kredit.....	59
4.5	Hasil Evaluasi Model pada Threshold Optimal	60
4.5.1	Alasan Penggunaan Threshold Optimal.....	60

4.5.2	Penentuan Threshold Optimal.....	61
4.5.3	Hasil Evaluasi Gaussian Naive Bayes pada Threshold Optimal ...	61
4.5.4	Hasil Evaluasi Random Forest pada Threshold Optimal	63
4.5.5	Analisis Perbandingan Berdasarkan Confusion Matrix	65
4.5.6	Analisis Perubahan Kinerja Dibandingkan Threshold Default.....	66
4.5.7	Implikasi terhadap Deteksi Risiko Kredit.....	67
4.6	Sintesis Perbandingan Gaussian Naive Bayes dan Random Forest	67
4.6.1	Perilaku Model pada Threshold Default	68
4.6.2	Dampak Optimasi Threshold terhadap Kinerja Model	68
4.6.3	Analisis Trade-off antara Model Sederhana dan Model Kompleks 68	
4.7	Pembahasan terhadap Tujuan Penelitian dan Implikasi.....	69
4.7.1	Ketercapaian Tujuan Penelitian	69
4.7.2	Implikasi Ketidakseimbangan Data dan Strategi Evaluasi	69
BAB V		71
PENUTUP		71
5.1	Kesimpulan	71
5.2	Saran	72
REFERENSI		74
LAMPIRAN		76

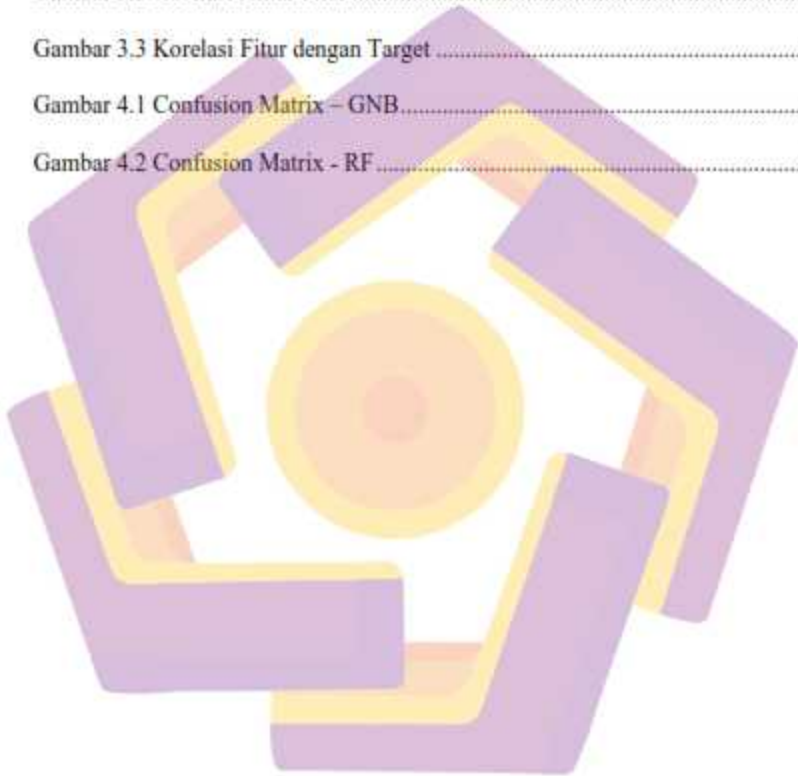
DAFTAR TABEL

Table 2.1 Keaslian Penelitian	12
Table 4.1 Daftar dan Deskripsi Fitur pada Dataset Akhir.....	48
Table 4.2 Distribusi Kelas Target loan_status	51
Table 4.3 Klasifikasi Tipe Fitur dan Implikasinya terhadap Pemodelan.....	53
Table 4.4 Hasil Evaluasi Model pada Threshold Default (0.5)	57
Table 4.5 Hasil Evaluasi Model pada Threshold Optimal	61



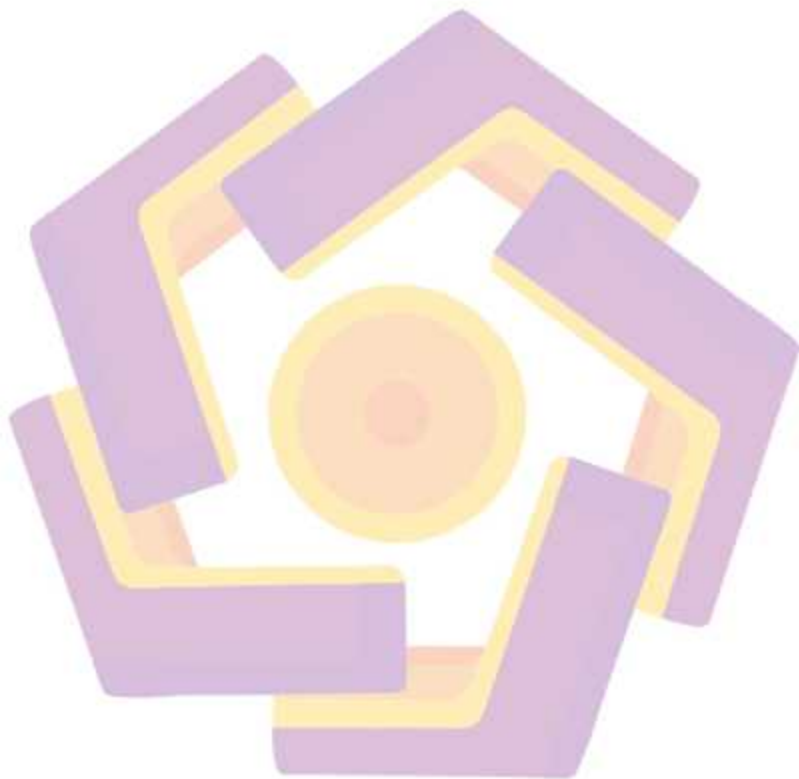
DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	35
Gambar 3.2 Korelasi Antar Fitur	38
Gambar 3.3 Korelasi Fitur dengan Target	39
Gambar 4.1 Confusion Matrix – GNB.....	62
Gambar 4.2 Confusion Matrix - RF.....	64



DAFTAR LAMPIRAN

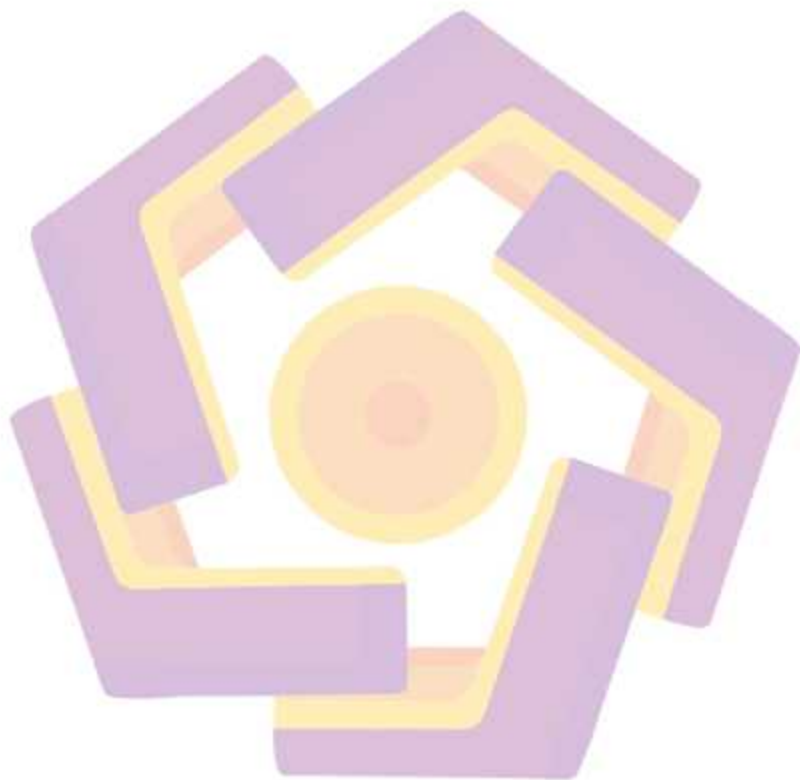
Lampiran 1 Dataset yg digunakan	76
---------------------------------------	----



DAFTAR LAMBANG DAN SINGKATAN

$P(x_i y)$	Probabilitas bersyarat dari nilai fitur ke- i (x_i) diberikan kelas y
x_i	Nilai observasi pada fitur ke- i dari suatu sampel
y	Label kelas (misalnya: $y = 0$ untuk non-default, $y = 1$ untuk default)
$\mu_{y,i}$	Rata-rata fitur ke- i pada kelas y , parameter lokasi distribusi Gaussian untuk kelas y
$\sigma_{y,i}$	Standar deviasi fitur ke- i pada kelas y , parameter skala distribusi Gaussian untuk kelas y
$\sigma_{y,i}^2$	Varians fitur ke- i pada kelas y
$\exp(\cdot)$	Fungsi eksponensial, yaitu $e^{(\cdot)}$, dengan $e = 2.71828$
Gini	Nilai indeks Gini — ukuran ketidakmurnian (impurity) suatu node dalam pohon keputusan
K	Jumlah total kelas dalam masalah klasifikasi (untuk kasus biner, $K = 2$)
p_k	Proporsi (frekuensi relatif) sampel dari kelas ke- k dalam suatu node atau subset data
$\sum_{k=1}^K$	Notasi sigma jumlah dari $k = 1$ hingga K
GNB	Gaussian Naive Bayes
P2P	Peer-to-Peer
RF	Random Forest
OJK	Otoritas Jasa Keuangan
TWP90	Tingkat Wanprestasi di atas 90 Hari
GIGO	Garbage In, Garbage Out
MAR	Missing At Random

MCAR	Missing Completely At Random
MNAR	Missing Not At Random
FICO	Fair Isaac Corporation



DAFTAR ISTILAH

At Application	Informasi yang digunakan dalam pemodelan hanya mencakup data yang tersedia pada saat pengajuan pinjaman.
Baseline Model	Model pembanding sederhana yang digunakan sebagai acuan awal untuk mengevaluasi peningkatan performa model lain.
Class Weight	Penyesuaian bobot pada kelas tertentu dalam proses pelatihan model untuk meningkatkan sensitivitas terhadap kelas minoritas.
Data Leakage	Kondisi ketika informasi yang seharusnya tidak tersedia pada saat prediksi digunakan dalam pelatihan model sehingga menghasilkan estimasi performa yang terlalu optimistis.
Out-of-Fold – Predictio	Prediksi yang dihasilkan dari data uji pada setiap fold dalam cross-validation, di mana observasi tidak dilatih pada model yang menghasilkan prediksi tersebut.
Pipeline	Rangkaian proses terstruktur yang menggabungkan tahapan pra-pemrosesan dan pelatihan model dalam satu alur sistematis.
Risk Signal	Indikator atau karakteristik dalam data yang mengandung informasi terkait kemungkinan terjadinya risiko gagal bayar.
Stratified Sampling	Teknik sampling yang mempertahankan proporsi kelas tertentu agar sesuai dengan distribusi populasi.
Threshold-Optimization	Proses penentuan nilai threshold yang menghasilkan performa evaluasi terbaik berdasarkan metrik tertentu.

Trade-off	Hubungan kompromi antara dua metrik evaluasi, seperti recall dan precision, di mana peningkatan satu metrik dapat menurunkan metrik lainnya.
Tree-Based Model	Model berbasis struktur pohon keputusan yang membagi data berdasarkan aturan tertentu.
Underfitting	Kondisi ketika model terlalu sederhana sehingga tidak mampu menangkap pola dalam data.
Overfitting	Kondisi ketika model terlalu kompleks dan terlalu menyesuaikan diri dengan data pelatihan sehingga performa pada data baru menurun.



INTISARI

Perkembangan layanan peer-to-peer (P2P) lending telah memperluas akses pembiayaan, namun secara simultan meningkatkan eksposur terhadap risiko kredit, khususnya risiko gagal bayar (default). Salah satu karakteristik utama data kredit pada platform P2P lending adalah ketidakseimbangan distribusi kelas antara peminjam lancar dan gagal bayar. Kondisi ini menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas sehingga menurunkan sensitivitas deteksi risiko dan berpotensi meningkatkan kerugian finansial apabila tidak ditangani dengan pendekatan evaluasi dan pengambilan keputusan yang tepat.

Hasil evaluasi menunjukkan bahwa peningkatan kompleksitas algoritma tidak secara otomatis meningkatkan performa pada threshold default (0,5). RF memiliki kemampuan diskriminatif probabilistik yang lebih baik dengan nilai ROC-AUC sebesar 0,7134 dibandingkan GNB sebesar 0,6468. Namun, pada threshold default, RF gagal mendeteksi kelas minoritas (Recall = 0,0309). Setelah dilakukan optimasi threshold, RF mencapai kinerja terbaik dengan F1-Score sebesar 0,4301 pada threshold 0,22, ditandai dengan peningkatan Recall yang substansial dari 0,0309 menjadi 0,6139 serta keseimbangan Precision-Recall yang lebih stabil. Sementara itu, GNB hanya mengalami peningkatan marginal pada F1-Score (dari 0,3842 menjadi 0,3884) pada threshold 0,63, dengan trade-off penurunan sensitivitas deteksi.

Temuan ini menegaskan bahwa pada data tidak seimbang tanpa modifikasi bobot kelas, efektivitas model kompleks sangat bergantung pada strategi penyesuaian threshold untuk mengelola trade-off risiko. Random Forest dengan optimasi threshold menunjukkan potensi yang lebih adaptif untuk diterapkan dalam sistem penilaian risiko P2P lending, karena mampu meningkatkan deteksi gagal bayar secara signifikan dengan tingkat kesalahan prediksi yang lebih terkendali dibandingkan model probabilistik sederhana.

Kata kunci: peer-to-peer lending, risiko kredit, Gaussian Naive Bayes, Random Forest, data tidak seimbang.

ABSTRACT

The development of peer-to-peer (P2P) lending services has expanded access to financing while simultaneously increasing exposure to credit risk, particularly default risk. One of the main characteristics of credit data on P2P lending platforms is class imbalance between performing and default borrowers. This condition causes classification models to be biased toward the majority class, thereby reducing risk detection sensitivity and potentially increasing financial losses if not addressed with appropriate evaluation and decision-making strategies.

The evaluation results indicate that increasing algorithmic complexity does not automatically improve performance at the default threshold (0.5). Random Forest (RF) demonstrates better probabilistic discriminative ability with a ROC-AUC of 0.7134 compared to 0.6468 for Gaussian Naive Bayes (GNB). However, at the default threshold, RF fails to adequately detect the minority class (Recall = 0.0309). After threshold optimization, RF achieves its best performance with an F1-Score of 0.4301 at a threshold of 0.22, marked by a substantial increase in Recall from 0.0309 to 0.6139 and a more stable Precision-Recall balance. In contrast, GNB shows only a marginal improvement in F1-Score (from 0.3842 to 0.3884) at a threshold of 0.63, accompanied by a trade-off in reduced detection sensitivity.

These findings confirm that in imbalanced data without class weight modification, the effectiveness of complex models strongly depends on threshold adjustment strategies to manage risk trade-offs. Random Forest with threshold optimization demonstrates greater adaptive potential for implementation in P2P lending risk assessment systems, as it significantly improves default detection while maintaining more controlled prediction errors compared to a simple probabilistic model.

Keyword: peer-to-peer lending, credit risk, Gaussian Naive Bayes, Random Forest, imbalanced data.