

**IMPLEMENTASI RANDOM FOREST UNTUK KLASIFIKASI
RISIKO PINJAMAN KEUANGAN**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ATILIA DHAFFAZRA HERMAWAN

21.11.4243

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

IMPLEMENTASI RANDOM FOREST UNTUK KLASIFIKASI RISIKO PINJAMAN KEUANGAN

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ATILIA DHAFFAZRA HERMAWAN

21.11.4243

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**IMPLEMENTASI RANDOM FOREST UNTUK KLASIFIKASI RISIKO
PINJAMAN KEUANGAN**

yang disusun dan diajukan oleh

Atilia Dhaffazra Hermawan

21.11.4243

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal <21 April 2025>

Dosen Pembimbing,



I Made Artha Agastya, S.T., M.Eng., Ph.D.

NIK. 190302352

HALAMAN PENGESAHAN
SKRIPSI
IMPLEMENTASI RANDOM FOREST UNTUK KLASIFIKASI RISIKO
PINJAMAN KEUANGAN

yang disusun dan diajukan oleh

Atilia Dhaffazra Hermawan

21.11.4243

Telah dipertahankan di depan Dewan Penguji
pada tanggal <19 Mei 2025>

Nama Penguji

Susunan Dewan Penguji

Tanda Tangan

Bambang Pilu Hartato, S.Kom., M.Eng
NIK. 190302707

Dina Maulina, M.Kom
NIK. 190302250

I Made Artha Agastya, S.T., M.Eng., Ph.D.
NIK. 190302352



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal <19 Mei 2025>

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Atilia Dhaffazra Hermawan
NIM : 21.11.4243

Menyatakan bahwa Skripsi dengan judul berikut:

Implementasi Random Forest untuk Klasifikasi Risiko Pinjaman Keuangan

Dosen Pembimbing : I Made Artha Agastya, S.T., M.Eng., Ph.D

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, <21 April 2025>

Yang Menyatakan,



Atilia Dhaffazra Hermawan

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa, atas segala rahmat dan karunia-Nya, saya persembahkan skripsi ini sebagai ungkapan terima kasih yang mendalam. Untuk kedua orang tua saya tercinta, yang selalu mendoakan, mendukung, dan memberikan kasih sayang serta semangat tanpa henti, saya dedikasikan hasil dari perjuangan ini. Kepada dosen pembimbing bapak I Made Artha Agastya, S.T., M.Eng., Ph.D yang telah sabar memberikan arahan, bimbingan, dan ilmu berharga selama proses penyusunan skripsi ini, saya sampaikan rasa hormat dan terima kasih yang sebesar-besarnya. Juga, kepada teman-teman saya yang selalu memberikan dukungan, semangat, dan kebersamaan yang berarti dalam setiap langkah perjalanan ini. Semoga karya ini dapat menjadi awal dari kontribusi saya dalam ilmu pengetahuan dan bermanfaat bagi banyak orang.

Yogyakarta, 19 April 2025

Penulis

DAFTAR ISI

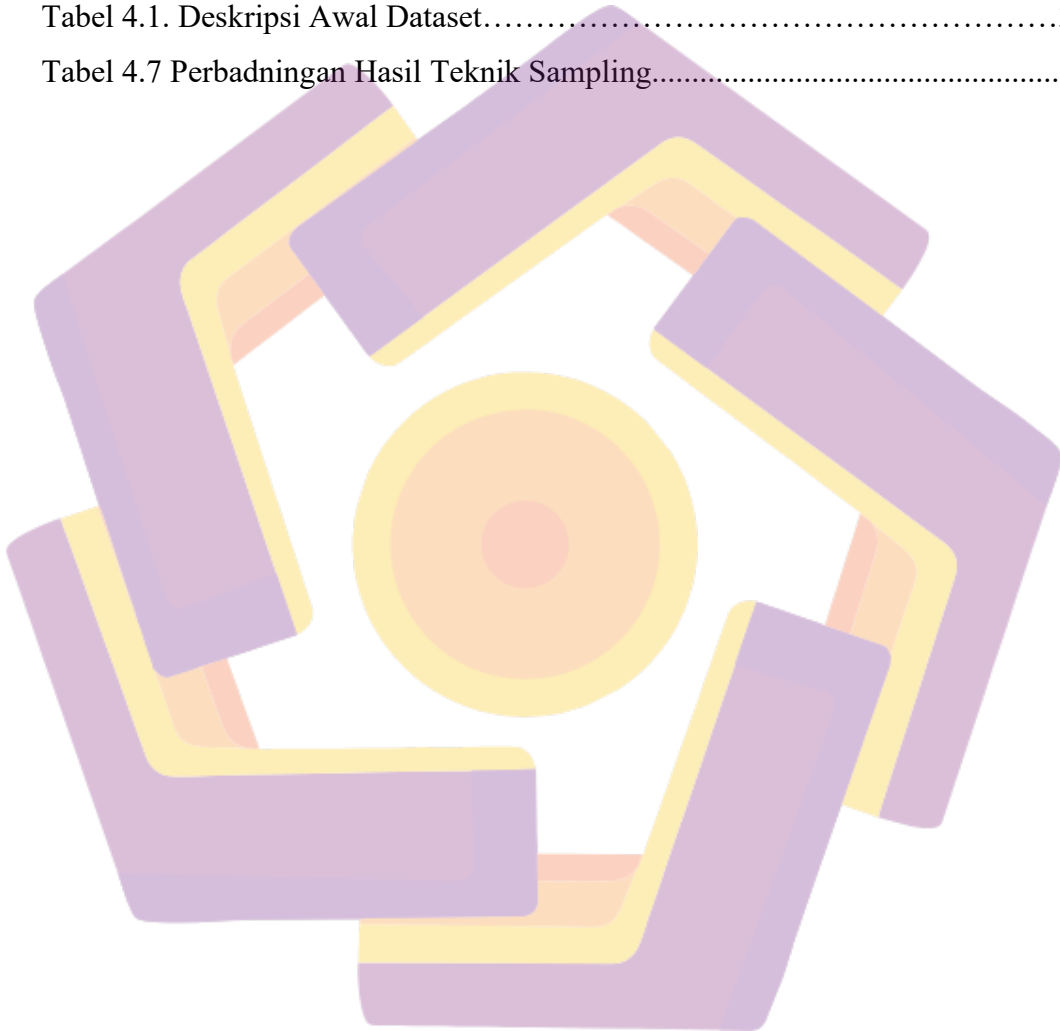
HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI.....	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur	6
Tabel 2.1.1 Perbandingan Algoritma	8

Tabel 2.1.2 Keaslian Penelitian.....	10
2.2 Dasar Teori.....	17
2.2.1 Ketidakseimbangan Data dalam Klasifikasi Risiko Pinjaman.....	17
2.2.2 Random Forest untuk Klasifikasi Risiko Pinjaman	19
2.2.3 Mengatasi Ketidakseimbangan Data dengan Teknik Oversampling ..	24
2.2.4 Evaluasi Kinerja Model dalam Klasifikasi Risiko Pinjaman.....	24
2.2.5 SMOTE dan Varian-Varian dalam Penanganan Ketidakseimbangan Data	25
2.2.6 Ensemble Learning dalam Prediksi Risiko Pinjaman	25
2.2.7 Pengolahan Data Keuangan dan Model Prediksi Risiko	25
2.2.8 Penggunaan SMOTE dalam Penanganan Ketidakseimbangan Data) ..	26
BAB III METODE PENELITIAN.....	27
3.1 Objek Penelitian	27
3.2 Alur Penelitian	27
3.2.1 Identifikasi dan Pemilihan Topik.....	28
3.2.2 Menetapkan Tujuan Penelitian.....	29
3.2.3 Studi Pustaka dan Pengumpulan Data.....	29
3.2.4 Pengumpulan Data dan Penanganan Ketidakseimbangan	29
3.2.5 Pengembangan Model dan Pemilihan Fitur	30
3.2.6 Pembagian Data dan Pelatihan Model	30
3.2.7 Evaluasi Model dan Pengujian.....	31
3.2.8 Penyempurnaan Model.....	31
3.2.9 Kesimpulan dan Saran	31
3.3 Alat dan Bahan.....	32
BAB IV HASIL DAN PEMBAHASAN	33

4.1 Data dan Persiapan Awal	33
4.2 Pemrosesan Data	34
4.3 Pembagian Data Latih dan Uji	34
4.4 Feature Importance and Selection Feature	34
4.5 Analisis Bivariat Karakteristik Pemohon Kredit terhadap Status Persetujuan Pinjaman.....	36
4.6 Evaluasi Performa Model: Perbandingan Teknik Sampling dan Penerapan KMeansSMOTE.....	64
4.7 Deployment.....	68
Proses Deployment	68
Tampilan Aplikasi.....	70
Hasil Prediksi “Loan Not Approved”	72
Hasil Prediksi “Loan Approved”	74
Pengujian Aplikasi	75
BAB V PENUTUP.....	77
5.1 Kesimpulan	77
5.2 Saran.....	78
REFERENSI	79
LAMPIRAN.....	83

DAFTAR TABEL

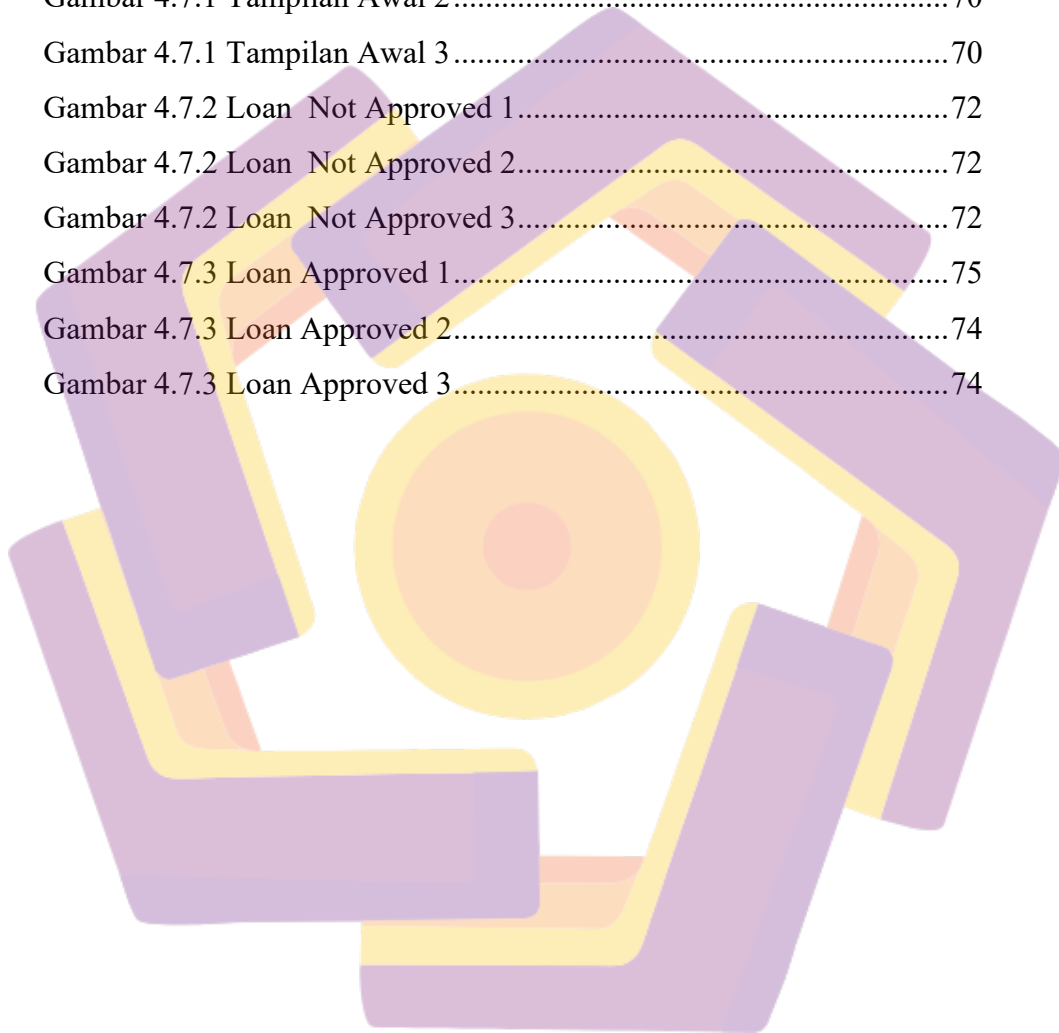
Tabel 2.1.1 Perbandingan Algoritma.....	8
Tabel 2.1.2 Keaslian Penelitian.....	10
Tabel 3.3. Library.....	32
Tabel 4.1. Deskripsi Awal Dataset.....	33
Tabel 4.7 Perbadningan Hasil Teknik Sampling.....	64



DAFTAR GAMBAR

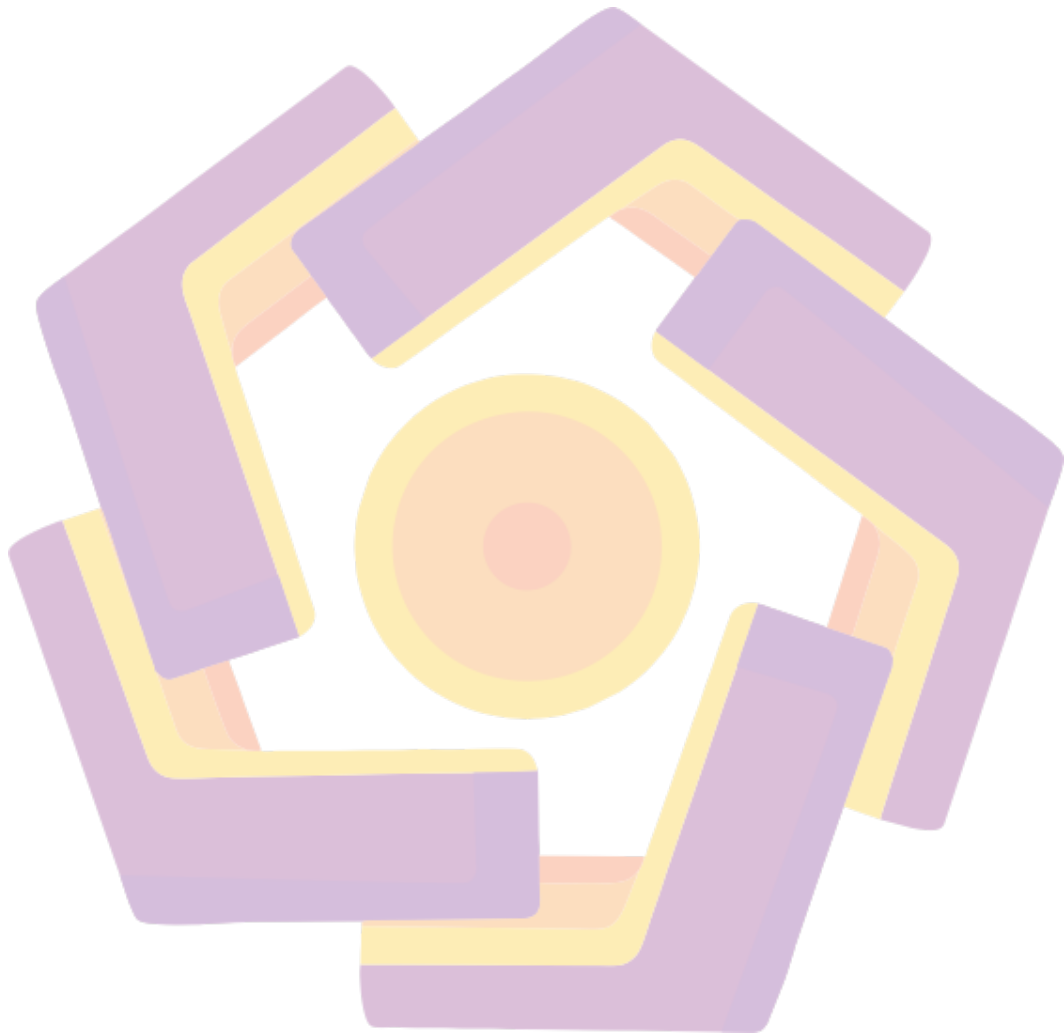
Gambar 2.2.2 Random Forest	22
Gambar 2.2.2 Flowchart Alur Penelitian	23
Gambar 3.2 Flowchart Alur Penelitian	28
Gambar 4.4 Feature Importance.....	35
Gambar 4.5 Loan Approved vs Annual Income	37
Gambar 4.5 Loan Approved vs Age	38
Gambar 4.5 Loan Approved vs Credit Score.....	39
Gambar 4.5 Loan Approved vs Education Level.....	40
Gambar 4.5 Loan Approved vs Experience	41
Gambar 4.5 Loan Approved vs Loan Amount.....	42
Gambar 4.5 Loan Approved vs Loan Duration.....	43
Gambar 4.5 Loan Approved vs Number Of Dependents.....	44
Gambar 4.5 Loan Approved vs Monthly Debt Payments	45
Gambar 4.5 Loan Approved vs Credit Card Utilization Rate.....	46
Gambar 4.5 Loan Approved vs Number Of Open Credit Lines	47
Gambar 4.5 Loan Approved vs Number Of Credit Inquiries.....	48
Gambar 4.5 Loan Approved vs Base Interest Rate	49
Gambar 4.5 Loan Approved vs Checking Account Balance	50
Gambar 4.5 Loan Approved vs Credit Score.....	51
Gambar 4.5 Loan Approved vs Debt To Income Ratio	52
Gambar 4.5 Loan Approved vs Interest Rate.....	53
Gambar 4.5 Loan Approved vs Job Tenure	54
Gambar 4.5 Loan Approved vs Length Of Credit History.....	55
Gambar 4.5 Loan Approved vs Monthly Income	56
Gambar 4.5 Loan Approved vs Monthly Loan Payment	57
Gambar 4.5 Loan Approved vs Net Worth	58
Gambar 4.5 Loan Approved vs Payment History	59
Gambar 4.5 Loan Approved vs Rick Score	60
Gambar 4.5 Loan Approved vs Total Assets	61

Gambar 4.5 Loan Approved vs Total Debt To Income Ratio.....	62
Gambar 4.5 Loan Approved vs Total Liabilities	63
Gambar 4.6 Standard Deviasi Instance Hardness threshold	66
Gambar 4.6 Confusion Matriks.....	67
Gambar 4.7.1 Tampilan Awal 1	70
Gambar 4.7.1 Tampilan Awal 2	70
Gambar 4.7.1 Tampilan Awal 3	70
Gambar 4.7.2 Loan Not Approved 1.....	72
Gambar 4.7.2 Loan Not Approved 2.....	72
Gambar 4.7.2 Loan Not Approved 3.....	72
Gambar 4.7.3 Loan Approved 1.....	75
Gambar 4.7.3 Loan Approved 2.....	74
Gambar 4.7.3 Loan Approved 3.....	74



DAFTAR LAMPIRAN

Lampiran kode.....	83
--------------------	----



DAFTAR LAMBANG DAN SINGKATAN

KDD
ROC
AUC
CRISP-DM

PSO
ANN
XGBoost:
SVM
IHT
JSON

CSS
HTML

Knowledge Discovery in Databases.
Receiver Operating Characteristics.
Area Under the Curve.
Cross-Industry Standard Process for
Data Mining.
Particle Swarm Optimization.
Artificial Neural Network.
Extreme Gradient Boosting.
Support Vector Machine.
Instance Hardness Threshold
JavaScript Object Notation (format
data yang digunakan dalam backend
aplikasi).
Cascading Style Sheets
HyperText Markup language

DAFTAR ISTILAH



Random Forest	Algoritma ensemble yang menggabungkan prediksi dari beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko overfitting.
Oversampling	Teknik untuk menyeimbangkan dataset dengan menambahkan data kelas minoritas.
Undersampling	Teknik untuk menyeimbangkan dataset dengan mengurangi data dari kelas mayoritas.
Feature Importance	Metode untuk mengukur kontribusi masing-masing fitur terhadap prediksi model.
Cross-Validation	Teknik evaluasi model yang membagi data menjadi beberapa subset untuk melatih dan menguji model secara bergantian.
Precision	Proporsi prediksi positif yang benar terhadap semua prediksi positif yang dihasilkan model.
Recall	Proporsi data positif yang benar-benar teridentifikasi oleh model.
F1-Score	Harmonic mean antara precision dan recall, digunakan untuk mengevaluasi kinerja model pada data yang tidak seimbang.
Akurasi	Rasio prediksi yang benar terhadap total prediksi yang dilakukan oleh model.
Loan Approved/Loan Failed	Status keputusan pinjaman berdasarkan model prediksi
Confusion Matrix	sebuah tabel yang digunakan untuk mengevaluasi kinerja dari model klasifikasi machine learning.

INTISARI

Perkembangan teknologi di sektor Financial Technology (Fintech) telah mendorong popularitas layanan pinjaman online. Namun, risiko kredit tetap menjadi tantangan utama, khususnya akibat ketidakseimbangan informasi antara peminjam dan pemberi pinjaman. Jumlah persetujuan pinjaman yang tinggi tidak selalu disertai dengan kemampuan membayar kembali, sehingga akurasi sistem prediksi risiko default menjadi rendah. Penelitian ini mengimplementasikan algoritma Random Forest untuk mengklasifikasikan risiko pinjaman, dengan pendekatan penanganan ketidakseimbangan data menggunakan teknik oversampling KMeansSMOTE. Tujuannya adalah meningkatkan sensitivitas model terhadap kelas minoritas, yaitu peminjam yang berisiko gagal bayar, tanpa mengorbankan akurasi secara keseluruhan. Hasil penelitian menunjukkan bahwa algoritma Random Forest efektif dalam mengklasifikasikan risiko pinjaman. Setelah penerapan KMeansSMOTE, model mencapai akurasi 99,43%, precision 99,51%, recall 99,03%, dan F1-Score 99,27%. Ini menunjukkan bahwa model mampu secara tepat mengidentifikasi peminjam yang layak maupun yang berisiko gagal bayar. Precision dan recall yang tinggi mencerminkan efektivitas model dalam mendeteksi kedua kategori tersebut, sementara F1-Score menandakan keseimbangan performa model secara keseluruhan. Seleksi fitur yang tepat juga membantu mengurangi kompleksitas model tanpa menurunkan akurasi. Kendala utama yang dihadapi adalah waktu komputasi yang cukup tinggi akibat penggunaan teknik oversampling. Untuk itu, penelitian lanjutan dapat difokuskan pada optimalisasi proses komputasi dan eksplorasi algoritma alternatif yang lebih efisien dalam menangani ketidakseimbangan data.

Kata kunci: Random Forest, KMeansSMOTE, Klasifikasi Risiko Pinjaman, Akurasi Prediksi, Ketidakseimbangan Data

ABSTRACT

The advancement of technology in the Financial Technology (Fintech) sector has driven the growing popularity of online lending services. However, credit risk remains a major challenge, particularly due to information asymmetry between borrowers and lenders. A high number of loan approvals does not always correspond to repayment capability, leading to low accuracy in default risk prediction systems. This study implements the Random Forest algorithm to classify loan risk, using the KMeansSMOTE oversampling technique to address data imbalance. The objective is to improve the model's sensitivity toward the minority class borrowers at risk of default without compromising overall accuracy. The results show that the Random Forest algorithm is effective in classifying loan risk. After applying KMeansSMOTE, the model achieved 99.43% accuracy, 99.51% precision, 99.03% recall, and a 99.27% F1-score. These figures indicate the model's strong ability to accurately identify both eligible borrowers and those at risk of default. High precision and recall reflect the model's effectiveness in detecting both classes, while the F1-score demonstrates balanced performance. Proper feature selection also helped reduce model complexity without sacrificing accuracy. The main limitation lies in the high computational time required by the oversampling technique. Future research can focus on optimizing computational efficiency and exploring alternative algorithms that are more effective in handling data imbalance.

Keyword: Random Forest, KMeansSMOTE, Loan Risk Classification, Prediction Accuracy, Imbalance Data