

**PERBANDINGAN ALGORITMA RANDOM FOREST
DAN LOGISTIC REGRESSION UNTUK PREDIKSI
RISIKO KREDIT BERDASARKAN DATA PINJAMAN
BANK MANDIRI**

SKRIPSI SARJANA SISTEM INFORMASI

Oleh

Adi Subhandi Subiyono

227006516087



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2025

**PERBANDINGAN ALGORITMA RANDOM FOREST
DAN LOGISTIC REGRESSION UNTUK PREDIKSI
RISIKO KREDIT BERDASARKAN DATA PINJAMAN
BANK MANDIRI**

SKRIPSI SARJANA

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem
Informasi dari Fakultas Teknologi Komunikasi dan Informatika

Oleh

Adi Subhandi Subiyono

227006516087



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2025

HALAMAN PENGESAHAN
TUGAS AKHIR

**PERBANDINGAN ALGORITMA RANDOM FOREST DAN
LOGISTIC REGRESSION UNTUK PREDIKSI RISIKO
KREDIT BERDASARKAN DATA PINJAMAN BANK**



Dosen Pembimbing

A handwritten signature in black ink, likely belonging to Prof. Dr. Agung Triayudi, is positioned above the printed name and ID number.

(Prof. Dr. Agung Triayudi, S.Kom., M.Kom)
NIDN. 0419068604

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul:

PERBANDINGAN ALGORITMA RANDOM FOREST DAN LOGISTIC REGRESSION UNTUK PREDIKSI RISIKO KREDIT BERDASARKAN DATA PINJAMAN BANK

MANDIRI

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian-bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 26 Februari 2026



Adi Subhandi Subiyono

227006516087

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul:

PERBANDINGAN ALGORITMA RANDOM FOREST DAN LOGISTIC REGRESSION UNTUK PREDIKSI RISIKO KREDIT BERDASARKAN DATA PINJAMAN BANK

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2025-2026 pada tanggal 24 Februari Tahun 2026

MANDIRI

Dosen Pembimbing

Prof. Dr. Agung Triayudi, S.Kom., M.Kom

NIDN. 0419068604

Ketua Program Studi

Dr. Ir. Andrianingsih, S.Kom., MMSI

NIDN. 0303097902

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Adi Subhandi Subiyono

NPM : 227006516087

Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika

Program Studi : Sistem Informasi

Tanggal Sidang : 24 Februari 2026

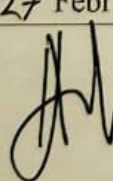
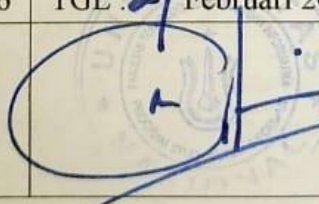
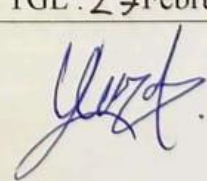
JUDUL DALAM BAHASA INDONESIA:

**PERBANDINGAN ALGORITMA RANDOM FOREST DAN
LOGISTIC REGRESSION UNTUK PREDIKSI RISIKO
KREDIT BERDASARKAN DATA PINJAMAN BANK
MANDIRI**

JUDUL DALAM BAHASA INGGRIS:

*COMPARISON OF RANDOM FOREST AND LOGISTIC
REGRESSION ALGORITHMS FOR CREDIT RISK PREDICTION
BASED ON BANK MANDIRI LOAN DATA*

TANDA TANGAN DAN TANGGAL

Dosen Pembimbing	Ka. Prodi	Mahasiswa
TGL : 27 Februari 2026	TGL : 27 Februari 2026	TGL : 27 Februari 2026
		

ABSTRAK

Risiko kredit merupakan salah satu faktor utama yang memengaruhi kualitas portofolio pembiayaan, sehingga dibutuhkan model prediksi yang tidak hanya akurat secara umum, tetapi juga andal dalam mendeteksi debitur berpotensi gagal bayar. Penelitian ini bertujuan membangun serta membandingkan kinerja model klasifikasi berbasis *machine learning* untuk memprediksi risiko kredit menggunakan data kredit Bank Mandiri, dengan target *default_flag* yang dibentuk berdasarkan kolektibilitas. Proses penelitian meliputi data preparation (pembersihan data, penanganan nilai hilang, transformasi variabel kategorikal melalui *one-hot encoding*, serta standarisasi fitur numerik), pembagian data latih dan uji secara stratified, serta penanganan ketidakseimbangan kelas menggunakan SMOTE pada data latih. Pemodelan dilakukan menggunakan *Logistic Regression* dan *Random Forest*, kemudian dioptimasi melalui *hyperparameter tuning* dengan cross-validation untuk memperoleh konfigurasi terbaik. Evaluasi model menekankan metrik yang relevan pada kelas minoritas, yaitu *Precision*, *Recall*, *F1-Score*, ROC-AUC, dan PR-AUC, sehingga hasil penilaian tidak bias terhadap dominasi kelas *non-default*. Hasil penelitian menunjukkan *Logistic Regression* (tuned) memberikan performa paling konsisten dalam mendeteksi kasus *default* dengan *Accuracy* 0,9894, *Precision* 0,8052, *Recall* 0,9688, *F1-Score* 0,8794, ROC-AUC 0,9975, PR-AUC 0,9307, serta menghasilkan *False Negative* 2 dan *False Positive* 15. Sementara itu, *Random Forest* (tuned) memperoleh *Accuracy* 0,9851, *Precision* 0,7564, *Recall* 0,9219, *F1-Score* 0,8310, ROC-AUC 0,9959, PR-AUC 0,8619, dengan *False Negative* 5 dan *False Positive* 19. Dengan demikian, *Logistic Regression* dinilai lebih efektif sebagai model early warning untuk mendukung mitigasi risiko kredit karena menekan peluang *default* yang terlewat sekaligus menjaga konsistensi pengambilan keputusan berbasis data.

Kata Kunci : Risiko kredit, Prediksi gagal bayar, Logistic Regression, Random Forest, SHAP.

ABSTRACT

Credit risk is a key factor influencing the quality of a financing portfolio, requiring a predictive model that is not only generally accurate but also reliable in detecting potential defaulters. This study aims to develop and compare the performance of a machine learning-based classification model for predicting credit risk using Bank Mandiri credit data, with a default_flag target formed based on collectibility. The research process included data preparation (data cleaning, handling missing values, transforming categorical variables through one-hot encoding, and standardizing numeric features), stratified data separation for training and testing, and handling class imbalance using SMOTE on the training data. Modeling was performed using Logistic Regression and Random Forest, then optimized through hyperparameter tuning with cross-validation to obtain the best configuration. Model evaluation emphasized metrics relevant to the minority class, namely Precision, Recall, F1-Score, ROC-AUC, and PR-AUC, to ensure that the assessment results were not biased towards the dominance of the non-default class. The results of the study showed that Logistic Regression (tuned) provided the most consistent performance in detecting default cases with Accuracy 0.9894, Precision 0.8052, Recall 0.9688, F1-Score 0.8794, ROC-AUC 0.9975, PR-AUC 0.9307, and produced 2 False Negatives and 15 False Positives. Meanwhile, Random Forest (tuned) obtained Accuracy 0.9851, Precision 0.7564, Recall 0.9219, F1-Score 0.8310, ROC-AUC 0.9959, PR-AUC 0.8619, with 5 False Negatives and 19 False Positives. Thus, Logistic Regression is considered more effective as an early warning model to support credit risk mitigation because it suppresses missed default opportunities while maintaining consistency in data-based decision making.

Keywords: Credit risk, Default prediction, Logistic Regression, Random Forest, SHAP.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunia sehingga penulis dapat menyelesaikan skripsi dengan judul **“PERBANDINGAN ALGORITMA RANDOM FOREST DAN LOGISTIC REGRESSION UNTUK PREDIKSI RISIKO KREDIT BERDASARKAN DATA PINJAMAN BANK MANDIRI”** sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional.

Penelitian dan penulisan skripsi ini tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan banyak terima kasih terutama kepada dosen pembimbing Tugas Akhir, Bapak Prof. Dr. Agung Triayudi, S.Kom., M.Kom. yang telah meluangkan banyak waktu, tenaga, pikiran, arahan, motivasi serta memaklumi segala kekurangan penulis selama penelitian tugas akhir dan penyusunan skripsi. Penulis juga mengucapkan banyak terima kasih kepada:

1. Allah Subhanahu Wa Ta'ala atas segala nikmat kesehatan, kekuatan, dan kemudahan yang diberikan selama proses penyelesaian skripsi ini.
2. Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Ayah, Ibu, Kakak, atas segala perhatian, doa, dan dukungan yang telah diberikan, yang tidak seluruhnya dapat penulis ungkapkan dengan kata-kata. Dukungan dari keluarga menjadi salah satu faktor utama yang menguatkan penulis hingga skripsi ini dapat diselesaikan
3. Bapak Prof. Dr. Agung Triayudi, S.Kom., M.Kom. selaku Dosen Pembimbing skripsi, yang telah membimbing saya untuk penulisan skripsi hingga selesai.
4. Ibu Dr. Aris Gunaryati, S.Si., MMSI. selaku Dosen Pembimbing Akademik, yang telah memberikan bimbingan dan arahan kepada penulis

sejak awal masa perkuliahan, khususnya dalam pemilihan mata kuliah, hingga penulis dapat menyelesaikan penyusunan skripsi ini.

5. Penulis juga menyampaikan terima kasih kepada teman-teman yang terlibat yang telah membantu dan mendukung penyelesaian skripsi ini, yang tidak dapat penulis sebutkan satu per satu. Penulis bersyukur dapat melalui proses penyusunan skripsi ini bersama kalian, senang bisa satu arc dengan kalian.
6. Penulis menyampaikan apresiasi kepada diri sendiri atas ketekunan, ketahanan, dan konsistensi selama proses penyusunan skripsi ini. Penulis juga bersyukur dapat mengambil pelajaran untuk menerima hal-hal yang berada di luar kendali, serta tetap menjaga komitmen hingga skripsi ini dapat diselesaikan.
7. *Last but not least, Let's fly higher*

Akhir kata, semoga Tuhan Yang Maha Esa membalas kebaikan dan bantuan yang telah diberikan dengan hal yang lebih baik. Penulis mengharapkan kritik dan saran yang bersifat membangun dan semoga skripsi ini dapat memberikan manfaat di bidang Teknologi Komunikasi dan Informatika.

Jakarta, 26 Februari 2026

Penulis



Adi Subhandi Subiyono
227006516087

DAFTAR ISI

HALAMAN PENGESAHAN TUGAS AKHIR.....	i
PERNYATAAN KEASLIAN TUGAS AKHIR.....	ii
LEMBAR PERSETUJUAN REVIEW AKHIR	iii
LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI	iv
KATA PENGANTAR.....	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	5
1.5 Kontribusi.....	6
BAB II TINJAUAN PUSTAKA	8
2.1 Studi Literatur	8
2.2 Landasan Teori	13
2.2.1 Risiko Kredit	13

2.2.2	Credit Scoring dan Komponen Risiko Kredit	13
2.2.3	Supervised Classification dalam Risiko Kredit.....	14
2.2.4	Logistic Regression.....	15
2.2.5	Random Forest	16
2.2.6	Ketidakseimbangan Kelas dan SMOTE.....	17
2.2.7	Evaluasi Model.....	18
2.2.8	Pra-pemrosesan data dan Hyperparameter Tuning	20
2.2.9	Explainable Artificial Intelligence (XAI).....	21
BAB III METODOLOGI PENELITIAN		23
3.1	Lokasi Penelitian.....	23
3.2	Waktu Penelitian	23
3.3	Subjek Penelitian.....	24
3.4	Fokus Penelitian.....	24
3.5	Sumber Data.....	25
3.6	Teknik Pengumpulan Data	27
3.7	Alur Penelitian	27
3.7.1	Data preparation.....	28
3.7.2	Preprocessing Data.....	29
3.7.3	Pembagian Data dan SMOTE	30
3.7.4	Modeling	31
3.7.5	Hyperparameter Tuning	34
3.7.6	Evaluasi Model.....	35
3.7.7	Simulasi Dampak Bisnis	37
3.7.8	Analisis dan Interpretasi Hasil	39
BAB IV HASIL DAN PEMBAHASAN.....		42

4.1	Data Preparation.....	42
4.2	Preprocessing Data.....	44
4.3	Pembagian Data dan SMOTE	46
4.4	Modeling	48
4.4.1	Random Forest	48
4.4.2	Logistic Regression.....	50
4.5	Hyperparameter Tuning	52
4.5.1	Random Forest	52
4.5.2	Logistic Regression.....	55
4.6	Evaluasi Model.....	57
4.7	Simulasi Dampak Bisnis	58
4.7.1	Hasil Simulasi Random Forest.....	60
4.7.2	Hasil Simulasi Logistic Regression	61
4.7.3	Perhitungan per Kontrak Random Forest.....	61
4.7.4	Perhitungan per Kontrak Logistic Regression	63
4.8	Analisis dan Interpretasi Hasil	64
4.8.1	Analisis Hasil	64
4.8.2	Interpretasi Hasil Model.....	66
BAB V KESIMPULAN DAN SARAN		72
5.1	Kesimpulan	72
5.2	Saran.....	73
DAFTAR PUSTAKA.....		74
LAMPIRAN.....		77

DAFTAR GAMBAR

Gambar 3. 1 Flowchart Alur Penelitian.....	27
Gambar 3. 2 Flowchart <i>Random Forest</i>	31
Gambar 3. 3 Flowchart <i>Logistic Regression</i>	33
Gambar 4. 1 Dataset awal	42
Gambar 4. 2 Variabel target.....	42
Gambar 4. 3 Penghapusan kolom	43
Gambar 4. 4 Pembentukan numerik & kategorikal.....	43
Gambar 4. 5 Ringkasan fitur & target.....	44
Gambar 4. 6 Missing dalam fitur kategorikal	44
Gambar 4. 7 Missing dalam fitur numerik	45
Gambar 4. 8 Pipeline preprocessing	45
Gambar 4. 9 Preview setelah preprocessing	46
Gambar 4. 10 Pembagian data	46
Gambar 4. 11 Distribusi sebelum & sesudah SMOTE.....	46
Gambar 4. 12 Confusion matrix RF.....	49
Gambar 4. 13 Heatmap confusion matrix RF	49
Gambar 4. 14 Confusion Matrix LR	51
Gambar 4. 15 Heatmap confusion matrix LR.....	52
Gambar 4. 16 Top kombinasi RF.....	53
Gambar 4. 17 Matrix setelah tuning RF.....	54
Gambar 4. 18 Heatmap matrix setelah tuning RF.....	54
Gambar 4. 19 Top kombinasi LR.....	55
Gambar 4. 20 Matrix setelah tuning LR.....	56
Gambar 4. 21 Heatmap matrix setelah tuning LR.....	56
Gambar 4. 22 SHAP global RF	66
Gambar 4. 23 Beeswarm RF	67
Gambar 4. 24 Highest risk RF.....	67
Gambar 4. 25 Lowest risk RF	68
Gambar 4. 26 SHAP global LR.....	69

Gambar 4. 27 Beeswarm LR.....	69
Gambar 4. 28 <i>True Positive</i> LR	70
Gambar 4. 29 <i>False Negative</i> LR.....	71



DAFTAR TABEL

Tabel 2. 1 Studi Literatur	8
Tabel 3. 1 Waktu Penelitian.....	23
Tabel 4. 1 Hasil Modeling RF	48
Tabel 4. 2 Hasil modeling LR	51
Tabel 4. 3 Hasil setelah tuning RF	53
Tabel 4. 4 Hasil setelah tuning LR	56
Tabel 4. 5 Ringkasan hasil RF & LR	58
Tabel 4. 6 Contoh Kontrak RF	62
Tabel 4. 7 Contoh Kontrak LR	63

