

**ANALISIS KOMPARATIF PERFORMA LOGISTIC
REGRESSION DAN RANDOM FOREST DALAM
KLASIFIKASI KELAYAKAN PENERIMAAN KREDIT**

SKRIPSI

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



disusun oleh
ANDREAS ADRIAN
21.11.4075

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**ANALISIS KOMPARATIF PERFORMA LOGISTIC
REGRESSION DAN RANDOM FOREST DALAM
KLASIFIKASI KELAYAKAN PENERIMAAN KREDIT**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ANDREAS ADRIAN

21.11.4075

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS KOMPARATIF PERFORMA LOGISTIC REGRESSION DAN
RANDOM FOREST DALAM KLASIFIKASI KELAYAKAN
PENERIMAAN KREDIT**

yang disusun dan diajukan oleh

Andreas Adrian

21.11.4075

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Januari 2025

Dosen Pembimbing,



Ike Verawati, S.Kom, M.Kom
NIK. 190302237

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS KOMPARATIF PERFORMA LOGISTIC REGRESSION DAN
RANDOM FOREST DALAM KLASIFIKASI KELAYAKAN
PENERIMAAN KREDIT**

yang disusun dan diajukan oleh

Andreas Adrian

21.11.4075

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Januari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Lilis Dwi Farida, S.Kom, M.Eng
NIK. 190302288

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

Ike Verawati, S.Kom, M.Kom
NIK. 190302237



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Januari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom, M.Kom, Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Andreas Adrian

NIM : 21.11.4075

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Komparatif Performa Logistic Regression dan Random Forest dalam Klasifikasi Kelayakan Penerimaan Kredit

Dosen Pembimbing : Ike Verawati, S.Kom, M.Kom

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, 22 Januari 2025

Yang Menyatakan,



Andreas Adrian

HALAMAN PERSEMBAHAN

Puji syukur kepada Tuhan Yang Maha Esa yang telah memberikan saya berkat dan anugrah sehingga dapat menyelesaikan skripsi ini dan memperoleh hasil yang terbaik. Skripsi ini saya persembahkan untuk :

1. Tuhan YME yang telah membimbing, menuntun, dan mengabulkan doa saya dalam menyelesaikan skripsi ini
2. Mending Nenek dan Tante Tercinta, yang telah berpulang ke hadirat Tuhan. Terima kasih atas kasih sayang, doa dan dukungan yang telah diberikan. Semoga mereka beristirahat dalam damai di sisi-Nya.
3. Kakek, Om, Ibu, Bapak, dan Kakak saya yang telah memberikan segala bentuk dukungan serta nasihat berharga bagi saya.
4. Ibu Ike Verawati, M.Kom selaku Dosen Pembimbing, saya sangat berterimakasih atas waktu, perhatian, nasihat serta masukan yang diberikan selama bimbingan hingga skripsi ini dapat terselesaikan dengan tepat waktu, baik dan lancar.
5. Bapak Ali Mustopa, M.Kom dan Ibu Lilis Dwi Farida M.Eng, selaku Dosen Penguji, yang telah memberikan saran dan masukan berharga dalam ujian pendadaran, sehingga skripsi ini dapat menjadi lebih baik.
6. Teman-teman seperjuangan, yang selalu saling support di waktu suka maupun duka.

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga skripsi ini dapat diselesaikan dengan baik sebagai salah satu syarat untuk memperoleh gelar Sarjana.

Penyusunan skripsi ini tidak lepas dari dukungan dan bantuan berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Ibu Ike Vrawati, S.Kom, M.Kom, selaku Dosen Pembimbing, atas bimbingan, arahan, serta masukan yang sangat berarti dalam penyusunan skripsi ini.
2. Bapak Ali Mustopa, S.Kom, M.Kom dan Ibu Lilis Dwi Farida, S.Kom, M.Eng, selaku Tim Dosen Penguji, atas saran dan kritik yang membangun dalam penyempurnaan skripsi ini.
3. Para Bapak/Ibu Dosen Universitas Amikom Yogyakarta, atas ilmu dan pengetahuan yang telah diberikan.
4. Kedua orang tua dan keluarga, atas segala doa, dukungan, serta motivasi yang tiada henti.
5. Rekan-rekan seperjuangan, yang telah memberikan semangat, bantuan, serta kebersamaan dalam menjalani proses perkuliahan hingga penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi kontribusi yang berarti dalam bidang keilmuan.

Yogyakarta, 22 Januari 2025

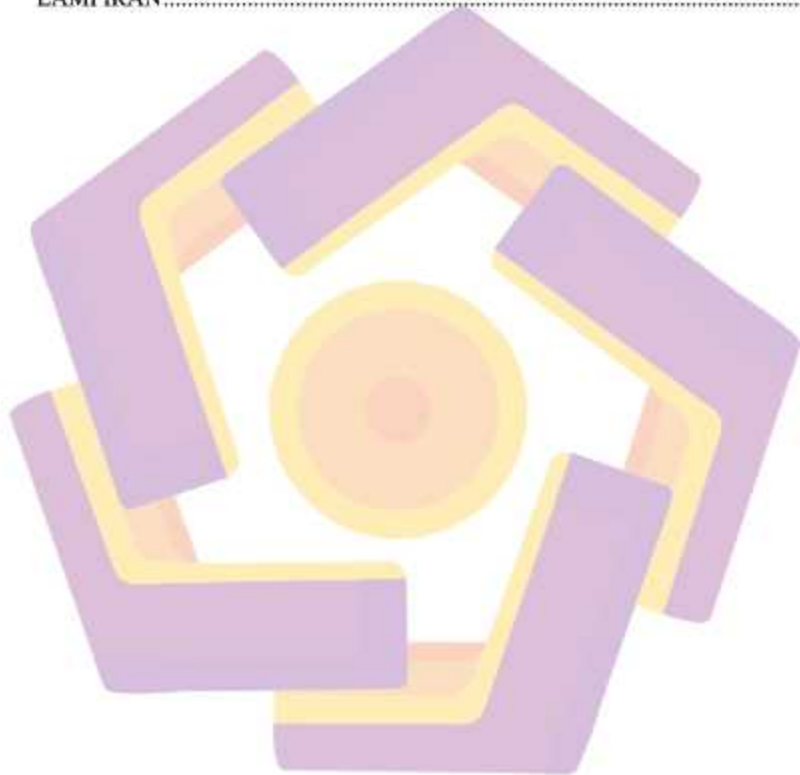
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
<i>ABSTRACT</i>	xv
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	5
2.1 Studi Literatur	5

2.2	Dasar Teori.....	10
2.2.1	Machine Learning	10
2.2.2	Dataset.....	10
2.2.3	Logistic Regression.....	10
2.2.4	Random Forest Classifier.....	12
2.2.5	Konsep Linear dan Non-Linear dalam dataset.....	12
2.2.6	Pre-Processing.....	13
2.2.7	Matriks Evaluasi Model	14
BAB III METODE PENELITIAN		16
3.1	Alur Penelitian.....	16
3.2	Alat dan Bahan.....	19
BAB IV HASIL DAN PEMBAHASAN		22
4.1	Akuisisi Data.....	22
4.2	EDA	23
4.3	Pre-Processing data.....	26
4.3.1	Data Standarization	26
4.3.2	Data Cleaning.....	29
4.3.3	Label Encoding	33
4.3.4	Balancing Data.....	38
4.3.5	Feature Selection.....	40
4.4	Modelling.....	59
4.4.1	Logistic Regression.....	61
4.4.2	Random Forest Classifier.....	62
4.5	Evaluasi.....	63
4.6	Analisis Evaluasi Model	71

BAB V PENUTUP	74
5.1 Kesimpulan	74
5.2 Saran	74
REFERENSI	75
LAMPIRAN	78



DAFTAR TABEL

Tabel 2.1 keaslian penelitian	7
Tabel 3.1 Sampel data	19
Tabel 3.2 Perangkat keras	20
Tabel 3.3 Perangkat lunak	20
Tabel 4.1 Sampel data	22
Tabel 4.2 Output head	24
Tabel 4.3 Output informasi dataset	25
Tabel 4.4 Hasil setelah penyeragaman format data	27
Tabel 4.5 Perbuahan yang telah dilakukan pada proses standarization	28
Tabel 4.6 Output informasi missing value	29
Tabel 4.7 Output dari <code>.value_counts()</code>	31
Tabel 4.8 Perbedaan jumlah missing value sebelum dan sesudah data cleaning	32
Tabel 4.9 Output dari informasi fitur pada dataset	33
Tabel 4.10 Output dari informasi fitur pada dataset	35
Tabel 4.11 Fitur "home_ownership" encoding	35
Tabel 4.12 Fitur "loan_intent" encoding	36
Tabel 4.13 Fitur "loan_grade" encoding	36
Tabel 4.14 Fitur "historical_default" encoding	37
Tabel 4.15 Fitur "current_loan_status" encoding	37
Tabel 4.16 Komposisi variabel y pada dataset	38
Tabel 4.17 Klasifikasi fitur berdasarkan karakteristik	57
Tabel 4.18 Output hasil evaluasi kedua model	64
Tabel 4.19 Waktu pelatihan model	72

DAFTAR GAMBAR

Gambar 2.1 Fungsi logit	10
Gambar 2.1 Hasil dari fungsi yang dihasilkan	11
Gambar 2.2 Ilustrasi Linear pada korelasi variabel	12
Gambar 2.3 Ilustrasi Non-Linear pada korelasi variabel	13
Gambar 3.1 Alur Penelitian	16
Gambar 4.1 Alur proses EDA	23
Gambar 4.2 Alur Proses Pre-Processing data	26
Gambar 4.3 Kondisi sebelum dan setelah balancing data	39
Gambar 4.4 Grafik heatmap	41
Gambar 4.5 Hasil korelasi antara fitur "customer_age" dengan y	43
Gambar 4.6 Hasil korelasi antara fitur "customer_income" dengan y	45
Gambar 4.7 Hasil korelasi antara fitur "employment_duration" dengan y	47
Gambar 4.8 Hasil korelasi antara fitur "loan_amnt" dengan y	49
Gambar 4.9 Hasil korelasi antara fitur "term_years" dengan y	51
Gambar 4.10 Hasil korelasi antara fitur "cred_hist_length" dengan y	53
Gambar 4.11 Hasil korelasi antara fitur "loan_intent_numeric" dengan y	55
Gambar 4.12 Hasil korelasi antara fitur "historical_default_numeric" dengan y	57
Gambar 4.13 Alur Proses Modelling	59
Gambar 4.14 Output dari proporsi dan pembagian data train dan test	60
Gambar 4.15 Output model dari hasil training model logistic regression	61
Gambar 4.16 Output model dari hasil training model Random Forest Classifier	62
Gambar 4.17 Alur Proses Evaluasi Model	63
Gambar 4.18 Confusion Matrix dari Logistic Regression	65
Gambar 4.19 Confusion Matrix dari Random Forest	66
Gambar 4.20 Grafik AUC-ROC dari Logistic Regression	68
Gambar 4.21 Grafik AUC-ROC dari Random Forest Classifier	70
Gambar 4.22 Alur Proses Analisis Evaluasi Model	71
Gambar 4.23 Perbandingan performa kedua model	71
Gambar 4.24 Grafik AUC-ROC kedua model	72

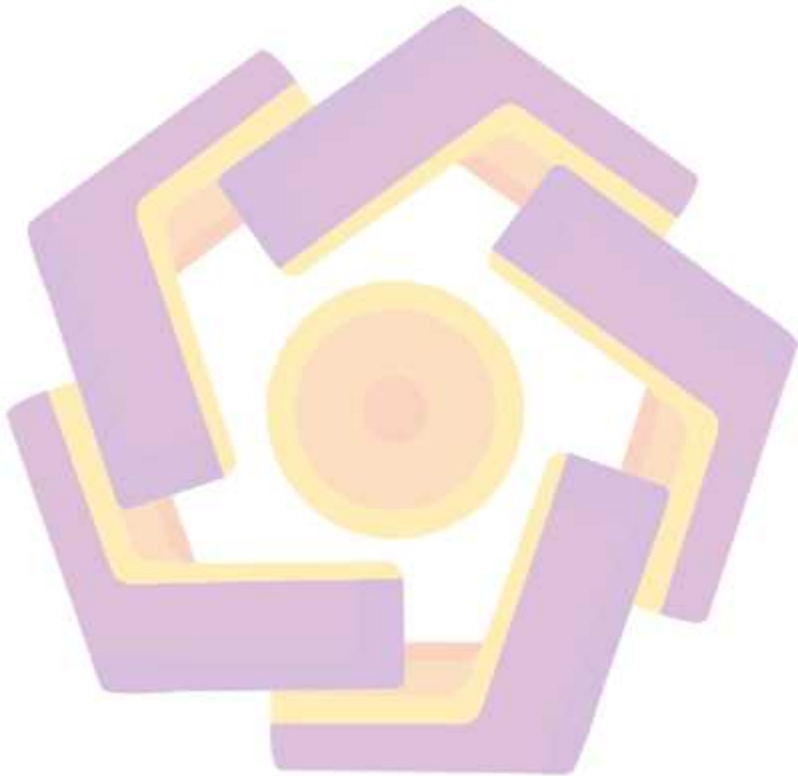
DAFTAR LAMBANG DAN SINGKATAN



LR	Logistic Regression
RF	Random Forest
KNN	K-Nearest Neighbor
SVM	Support Vector Machine
SGD	Stochastic Gradient Descent
UMKM	Usaha Mikro Kecil dan Menengah
SOP	Standar Operasional Prosedur
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
TN	True Negative
TP	True Positive
FN	False Negative
FP	False Positive
EDA	Exploratory Data Analysis

DAFTAR ISTILAH

Model	Algoritma machine learning
Variabel x	Variabel independen
Variabel y	Variabel dependen



INTISARI

Penentuan kelayakan penerimaan kredit merupakan proses yang sangat penting dalam industri perbankan dan keuangan. Untuk meningkatkan akurasi dan efisiensi dalam klasifikasi kelayakan kredit, penelitian ini berfokus pada penerapan dua model machine learning, yaitu Logistic Regression dan Random Forest Classifier. Logistic Regression dipilih karena kemampuannya dalam mengidentifikasi hubungan linear antara variabel input dan output, sedangkan Random Forest Classifier memiliki keunggulan dalam menangani dataset yang kompleks dan non-linear. Tujuan utama dari penelitian ini adalah untuk membandingkan performa kedua model tersebut dalam tugas klasifikasi kelayakan kredit. Perbandingan dilakukan dengan tahapan Studi Literatur, Akuisisi Data, EDA, Pre-Processing, Modelling, Evaluasi, dan Analisis Evaluasi Model. Dataset yang akan digunakan mencakup informasi demografis serta data finansial dari nasabah. Perbandingan performa pada penelitian ini menggunakan matrix akurasi, precision, recall, F1-Score dan AUC-ROC untuk mengevaluasi kinerja masing-masing model. Penelitian ini menghasilkan bahwa model random forest lebih unggul dengan skor Akurasi 0.85, Presisi 0.81, Recall 0.91 dan F1 Score 0.86. Skor AUC yang digunakan untuk melihat seberapa baik model dalam membedakan class 1 dan 0 mencapai 0.92. Hasil penelitian ini diharapkan mampu memberikan rekomendasi yang bermanfaat bagi industri perbankan dalam memilih model yang paling tepat untuk penilaian kelayakan kredit.

Kata kunci: Machine Learning, Kelayakan Kredit, Logistic Regression, Random Forest Classifier, Klasifikasi.

ABSTRACT

The determination of credit eligibility is a vital process in the banking and financial industries. To enhance accuracy and efficiency in credit eligibility classification, this study focuses on the application of two machine learning models: Logistic Regression and Random Forest Classifier. Logistic Regression was chosen for its ability to identify linear relationships between input and output variables, while Random Forest Classifier is advantageous in handling complex and non-linear datasets. The main objective of this study is to compare the performance of these two models in credit eligibility classification tasks. The comparison involves several stages: Literature Review, Data Acquisition, Exploratory Data Analysis (EDA), Pre-Processing, Modeling, Evaluation, and Model Evaluation Analysis. The dataset used includes demographic and financial information of customers. The performance comparison in this study utilizes accuracy, precision, recall, F1-Score, and AUC-ROC metrics to evaluate the effectiveness of each model. The study concludes that the Random Forest model outperforms Logistic Regression, achieving an accuracy score of 0.85, precision of 0.81, recall of 0.91, and F1-Score of 0.86. The AUC score, which measures the model's ability to distinguish between class 1 and 0, reached 0.92. The findings of this study are expected to provide valuable recommendations for the banking industry in selecting the most suitable model for credit eligibility assessment.

Keyword: Machine Learning, Credit Eligibility, Logistic Regression, Random Forest Classifier, Classification.