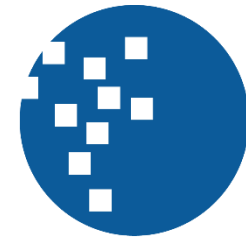


**KOMPARASI PERFORMA ALGORITMA *MACHINE*
LEARNING DALAM MENGANALISIS *NON-PERFORMING*
LOAN PADA INDUSTRI PERBANKAN**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Kezia Cheryl Sitinjak

00000061410

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

**KOMPARASI PERFORMA ALGORITMA *MACHINE*
LEARNING DALAM MENGANALISIS *NON-PERFORMING*
LOAN PADA INDUSTRI PERBANKAN**



SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer (S.Kom)

Kezia Cheryl Sitinjak
00000061410

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : **Kezia Cheryl Sitinjak**

Nomor Induk Mahasiswa : **00000061410**

Program Studi : **Sistem Informasi**

Skripsi dengan judul:

**KOMPARASI PERFORMA ALGORITMA *MACHINE LEARNING* DALAM
MENGANALISIS *NON-PERFORMING LOAN* PADA INDUSTRI
PERBANKAN**

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 16 Mei 2025



(Kezia Cheryl Sitinjak)

HALAMAN PENGESAHAN

Skripsi dengan judul

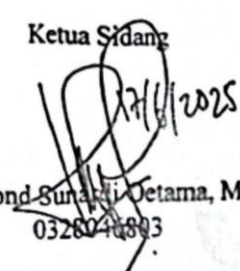
**KOMPARASI PERFORMA ALGORITMA *MACHINE LEARNING* DALAM
MENGANALISIS *NON-PERFORMING LOAN* PADA INDUSTRI
PERBANKAN**

Oleh

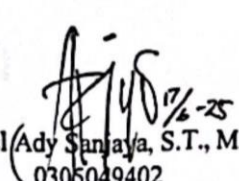
Nama : Kezia Cheryl Sitinjak
NIM : 00000061410
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Rabu, 4 Mei 2025
Pukul 15.00 s.d 17.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut.

Ketua Sidang


Ir. Raymond Sunardi Oetama, M.CIS.
0328046893

Penguji


Samuel Ady Sanjaya, S.T., M.T.
0306049402

Pembimbing


Prof. Friska Natalia, S.Kom., M.T.
0306128307

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom.
03130588001

iv

Komparasi Performa Algoritma..., Kezia Cheryl Sitinjak, Universitas Multimedia Nusantara

iii

Komparasi Performa Algoritma..., Kezia Cheryl Sitinjak, Universitas Multimedia Nusantara

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Kezia Cheryl Sitinjak

NIM : 00000061410

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : *Komparasi Performa Algoritma Machine Learning Dalam Menganalisis Non-Performing Loan Pada Industri Perbankan*

Menyatakan dengan sesungguhnya bahwa saya bersedia* (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 16 Mei 2025



(Kezia Cheryl Sitinjak)

* Pilih salah satu

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji Syukur atas selesainya penulisan Laporan Magang ini dengan judul: “KOMPARASI PERFORMA ALGORITMA *MACHINE LEARNING* DALAM MENGANALISIS *NON-PERFORMING LOAN* PADA INDUSTRI PERBANKAN” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Jurusan Sistem Informasi Pada Fakultas Teknik dan Informasika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Dr. Andrey Andoko selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Prof. Dr. Friska Natalia, S.Kom., M.T., sebagai dosen pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Industri perbankan yang telah memberikan bantuan dan dukungan terhadap data penelitian skripsi.
6. Orang tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
7. Timothy sebagai pasangan saya dan teman-teman saya yang telah memberikan semangat dan dukungan selama proses penyusunan tugas akhir ini.

Semoga karya ilmiah ini dapat bermanfaat bagi pembaca dan dapat dijadikan referensi untuk laporan Skripsi selanjutnya.

Tangerang, 16 Mei 2025



(Kezia Cheryl Sitinjak)

KOMPARASI PERFORMA ALGORITMA *MACHINE* *LEARNING* DALAM MENGANALISIS *NON-PERFORMING* *LOAN* PADA INDUSTRI PERBANKAN

Kezia Cheryl Sitinjak

ABSTRAK

Non-Performing Loan (NPL) merupakan kredit bermasalah yang tidak dapat dilunasi oleh para nasabah sesuai dengan perjanjian atau ketentuan yang telah disepakati bersama yang biasanya. Tingginya NPL dapat menyebabkan berbagai jenis kerugian pada sektor perbankan, seperti kehilangan pendapatan bunga dari pinjaman yang tidak dibayar oleh nasabah dan berdampak terhadap kepercayaan investor, pemegang saham, serta nasabah. Penelitian terdahulu menunjukkan bahwa model algoritma *XGBoost* dan *Random Forest* memberikan hasil klasifikasi yang tinggi. Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis NPL dengan membandingkan performa dari beberapa algoritma *machine learning*.

Algoritma *machine learning* yang digunakan pada penelitian ini adalah *Random Forest*, *Decision Tree*, dan *XGBoost*. Dataset yang digunakan terdiri dari 1.023.642 baris yang terdiri dari 818.913 untuk data train dan 204.729 untuk data test, serta terdiri dari 4 kolom yang telah dipilih melalui teknik *feature importance*. Data yang digunakan menggunakan keseluruhan *dataset* yang berisikan kelima kategori kredit. Untuk meningkatkan akurasi maka dilakukan *tuning* menggunakan *GridSearchCV*. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *f1-score*, dan *confusion matrix*.

Hasil evaluasi menunjukkan bahwa model *XGBoost* memberikan hasil kinerja yang paling unggul dengan *accuracy* mencapai 96%, *precision* 94%, *recall* 88%, dan *f1-score* 91%, menjadikannya model yang paling efektif dalam mengklasifikasi NPL. Penelitian ini membuktikan dan menyarankan penggunaan algoritma *XGBoost* sebagai algoritma utama untuk menganalisis dan memprediksi NPL karena *XGBoost* memiliki kemampuan yang tinggi dalam memberikan akurasi yang lebih tinggi dan lebih stabil dalam mengidentifikasi kelas.

Kata Kunci: *Decision Tree*, *Machine Learning*, *Non-Performing Loan* (NPL), *Random Forest*, *XGBoost*.

Comparison of Machine Learning Algorithm Performance in Analyzing Non-Performing Loan at Banking Industry

Kezia Cheryl Sitinjak

ABSTRACT (English)

Non-Performing Loan (NPL) refers to problematic loans that customers are unable to repay according to the terms or agreements that have been mutually agreed upon. A high level of NPL can cause various types of losses in the banking sector, such as the loss of interest income from loans that customers fail to repay, and it can negatively impact investors, shareholders, and customer confidence. Previous studies have shown that the XGBoost and Random Forest algorithms provide high classification results. Based on this, the purpose of this study is to analyze NPL by comparing the performance of several machine learning algorithms.

The machine learning algorithms used in this study are Random Forest, Decision Tree, and XGBoost. The dataset used consists of 1,023,642 rows, with 818,913 rows for the training data and 204,729 rows for the test data, as well as 4 columns selected through feature importance techniques. The data used includes the entire dataset containing the five credit categories. To improve accuracy, tuning was performed using GridSearchCV. Model evaluation was conducted using the metrics of accuracy, precision, recall, F1-score, and confusion matrix.

The evaluation results show that the XGBoost model delivers the best performance, achieving an accuracy of 96%, precision of 94%, recall of 88%, and F1-score of 91%, making it the most effective model for classifying NPL. This study proves and recommends the use of the XGBoost algorithm as the main algorithm for analyzing and predicting NPL due to its high ability to provide higher and more stable accuracy in identifying classes.

Keywords: *Decision Tree, Machine Learning, Non-Performing Loan (NPL) Random Forest, XGBoost.*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
<i>ABSTRACT</i> (English).....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan dan Manfaat Penelitian.....	4
1.4.1 Tujuan Penelitian.....	4
1.4.2 Manfaat Penelitian	5
1.5 Sistematika Penulisan	5
BAB II LANDASAN TEORI	7
2.1 Penelitian Terdahulu	7
2.2 Teori Penelitian	9
2.2.1 <i>Non-Performing Loan (NPL)</i>	9
2.2.2 <i>Data Mining</i>	10
2.2.3 <i>Machine Learning</i>	11
2.2.4 <i>Supervised Learning</i>	12
2.2.5 <i>Ensemble Learning</i>	12
2.2.6 <i>Encoding</i>	13
2.2.7 <i>Log-Transform</i>	13
2.2.8 <i>Cross-Validation</i>	13
2.2.9 <i>GridSearchCV</i>	14

2.2.10	<i>Confusion Matrix</i>	14
2.3	Framework dan Algoritma Penelitian	16
2.3.1	<i>Random Forest</i>	16
2.3.2	<i>Decision Tree</i>	17
2.3.3	<i>XGBoost</i>	18
2.3.4	CRISP-DM	19
2.4	Tools dan Software Penelitian	20
2.4.1	<i>Visual Studio Code</i>	20
2.4.2	<i>Python</i>	21
2.4.3	Microsoft Excel	21
2.4.4	<i>Streamlit</i>	21
BAB III	METODOLOGI PENELITIAN	23
3.1	Gambaran Umum Objek Penelitian	23
3.2	Metode Penelitian	23
3.3	Teknik Pengumpulan Data	26
3.3.1	Populasi dan Sampel	26
3.3.2	Periode Pengambilan Data	26
3.4	Teknik Analisis Data	26
3.5	Teknik Pengujian	28
3.6	Metode Pengolahan Data	29
BAB IV	ANALISIS DAN HASIL PENELITIAN	34
4.1	Business Understanding	34
4.2	Data Understanding	35
4.2.1	Deskripsi Data	36
4.2.2	<i>Exploratory Data Analysis (EDA)</i>	37
4.3	Data Preparation	46
4.3.1	<i>Data Cleaning</i> Pada Bulan Januari - Desember	46
4.3.2	Gabung Data Dari Bulan Januari – Desember	48
4.3.3	Konversi Tipe Data	49
4.3.4	Penghapusan Kolom	50
4.3.5	<i>Data Encoding</i>	50
4.3.6	Penanganan <i>Outlier</i>	51
4.3.7	Penanganan <i>Missing Values</i>	52
4.3.8	Penanganan Data Duplikat	53

4.3.9 Korelasi <i>Heatmap</i>	54
4.3.10 <i>Data Splitting</i>	55
4.3.11 <i>Feature Importance</i>	56
4.4 Modelling.....	59
4.4.1 Algoritma <i>Random Forest</i>	59
4.4.2 Algoritma <i>Decision Tree</i>	61
4.4.3 Algoritma <i>XGBoost</i>	63
4.5 Evaluation	66
4.6 Deployment	72
4.7 Hasil Penelitian	74
4.8 Pembahasan Hasil Penelitian	76
BAB V SIMPULAN DAN SARAN	78
5.1 Simpulan	78
5.2 Saran	78
DAFTAR PUSTAKA	80
LAMPIRAN.....	86



DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	7
Tabel 3. 1 Perbandingan Bahasa Pemrograman.....	26
Tabel 3. 2 Perbandingan <i>Tools Data Mining</i>	27
Tabel 3. 3 Perbandingan Teknik <i>Data Mining</i>	29
Tabel 3. 4 Perbandingan Algoritma Klasifikasi.....	32
Tabel 4. 1 Deskripsi Kolom pada <i>Dataset</i>	37
Tabel 4. 2 Jumlah Data yang Digunakan	37
Tabel 4. 3 Perbandingan <i>Cross-Validation</i> Setiap Algoritma.....	69
Tabel 4. 4 Perbandingan Kinerja Setiap Model <i>Machine Learning</i>	70
Tabel 4. 5 Perbandingan Performa Model <i>Machine Learning</i>	76



DAFTAR GAMBAR

Gambar 1. 1 Perkembangan Kredit dan NPL Bank Umum	2
Gambar 1. 2 Tabel Kinerja Industri Perbankan	3
Gambar 2. 1 Tabel <i>Confusion Matrix</i>	14
Gambar 2. 2 Cara Kerja <i>Random Forest</i>	16
Gambar 2. 3 Cara Kerja <i>Decision Tree</i>	18
Gambar 2. 4 Tahapan Metode CRISP-DM	19
Gambar 2. 5 Integrasi Anaconda.....	21
Gambar 3. 1 Kerangka Pemikiran	25
Gambar 3. 2 Alur Penelitian CRISP-DM.....	30
Gambar 4. 1 Surat Izin Pengambilan Data Penelitian (Skripsi).....	36
Gambar 4. 2 Struktur Dataset.....	36
Gambar 4. 3 Statistika Deskriptif.....	38
Gambar 4. 4 Distribusi Kolom Aging Tunggakan	39
Gambar 4. 5 Distribusi Kolom Baki Debet.....	40
Gambar 4. 6 Distribusi Kolom Nilai Tercatat.....	40
Gambar 4. 7 Distribusi Kolom Nilai Plafon	41
Gambar 4. 8 Distribusi Kolom Jumlah Tunggakan	41
Gambar 4. 9 Distribusi Aging Tunggakan Pada Kanwil	42
Gambar 4. 10 Distribusi Aging Tunggakan Berdasarkan 5 Produk Pinjaman Terbanyak.....	43
Gambar 4. 11 Distrbusi Persentase Baki Debet per Kanwil	43
Gambar 4. 12 Karakteristik Berdasarkan Aging Tunggakan	44
Gambar 4. 13 Boxplot Nilai Tercatat, Baki Debet, Plafon, Dan Jumlah Tunggakan	45
Gambar 4. 14 Boxplot NILAI_TERCATAT, BAKI_DEBET,	46
Gambar 4. 15 <i>Label Encoding</i>	47
Gambar 4. 16 Menggabungkan Data Bulanan	48
Gambar 4. 17 Menyimpan File Gabungan.....	49
Gambar 4. 18 Konversi Tipe Data Menjadi Numerik.....	49
Gambar 4. 19 <i>Data Encoding</i>	50
Gambar 4. 20 <i>Skewness</i> Pada Kolom.....	51
Gambar 4. 21 <i>Log Transformasi</i>	52
Gambar 4. 22 Hasil <i>Log Transformasi</i>	52
Gambar 4. 23 <i>Missing Values</i>	53
Gambar 4. 24 <i>Data Duplicate</i>	54
Gambar 4. 25 Korelasi <i>Heatmap</i>	55
Gambar 4. 26 <i>Data Splitting</i>	56
Gambar 4. 27 Hasil <i>Feature Importance</i>	57
Gambar 4. 28 Hasil <i>Feature Importance</i>	57
Gambar 4. 29 Fitur Yang Akan Digunakan	58
Gambar 4. 30 Hasil Akurasi <i>Training</i> dan <i>Test</i>	59
Gambar 4. 31 <i>Cross Validation Random Forest</i>	60
Gambar 4. 32 <i>Random Forest Accuracy</i>	61

Gambar 4. 33 Model <i>Decision Tree</i>	62
Gambar 4. 34 <i>Decision Tree Accuracy</i>	63
Gambar 4. 35 Model <i>XGBoost</i>	64
Gambar 4. 36 <i>GridSearchCV</i> Untuk Menemukan <i>Hyperparameter</i>	65
Gambar 4. 37 Cross-Validation dan Akurasi Pada <i>XGBoost</i>	66
Gambar 4. 38 <i>Confusion Matrix Random Forest</i>	67
Gambar 4. 39 <i>Confusion Matrix Decision Tree</i>	68
Gambar 4. 40 <i>Confusion Matrix XGBoost</i>	69
Gambar 4. 41 Perbandingan Metrik <i>Classification Report</i>	72
Gambar 4. 42 <i>Export Streamlit</i>	72
Gambar 4. 43 Hasil <i>Export</i>	72
Gambar 4. 44 <i>Code Streamlit</i>	73
Gambar 4. 45 <i>Output Streamlit</i>	74



DAFTAR RUMUS

Rumus 2. 1 Rumus <i>Accuracy</i> [29]	15
Rumus 2. 2 Rumus <i>Precision</i> [29]	15
Rumus 2. 3 Rumus <i>Recall</i> [29]	15
Rumus 2. 4 Rumus <i>F1-Score</i> [29]	16



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Turnitin <i>Similarity Report</i>	86
Lampiran B Form Konsultasi Bimbingan	87



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA