

**DETEKSI KECURANGAN KEUANGAN DENGAN
MENGUNAKAN METODE RANDOM FOREST DAN
XGBOOST PADA DATA YANG TIDAKSEIMBANG**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh
MUHAMMAD ALGHIFARI AL ASQOLANI
20.83.0564

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM
YOGYAKARTA YOGYAKARTA
2026

**DETEKSI KECURANGAN KEUANGAN DENGAN
MENGUNAKAN METODE RANDOM FOREST DAN
XGBOOST PADA DATA YANG TIDAKSEIMBANG
SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh
MUHAMMAD ALGHIFARI AL ASQOLANI
20.83.0564

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM
YOGYAKARTA YOGYAKARTA
2026

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

SKRIPSI

DETEKSI KECURANGAN KEUANGAN DENGAN MENGGUNAKAN METODE RANDOM FOREST DAN XGBOOST PADA DATA YANG TIDAKSEIMBANG

yang disusun dan diajukan oleh

Muhammad Alghifari Al Asqolani

20.83.0564

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 16 Desember 2025

Dosen Pembimbing,



Robert Mars, M. T., Ph.D
NIK. 190302228

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

SKRIPSI

**DETEKSI KECURANGAN KEUANGAN DENGAN MENGGUNAKAN
METODE RANDOM FOREST DAN XGBOOST PADA DATA YANG
TIDAKSEIMBANG**

yang disusun dan diajukan oleh

Muhammad Alghifari Al Asqolani

20.83.0564

Telah dipertahankan di depan Dewan Penguji
pada tanggal 16 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Jeki Kuswanto, S.Kom., M.Kom.
NIK. 190302456

Muhammad Kopravi, S.Kom., M.Eng.
NIK. 190302454

Robert Marco, M. T., Ph.D
NIK. 190302228



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 16 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Muhammad Alghifari Al Asqolani**
NIM : **20.83.0564**

Menyatakan bahwa Skripsi dengan judul berikut:

DETEKSI KECURANGAN KEUANGAN DENGAN MENGGUNAKAN METODE RANDOM FOREST DAN XGBOOST PADA DATA YANG TIDAKSEIMBANG

Dosen Pembimbing : Robert Marco, M. T., 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, 16 Desember 2025

Yang Menyatakan,



Muhammad Alghifari Al Asqolani

HALAMAN PERSEMBAHAN

Puji syukur penulis panjatkan kehadiran Allah SWT atas limpahan yang telah memberikan Kesehatan rohani dan jasmani, rahmat dan hidayah, sehingga penulis masih diberikan kesempatan untuk menyelesaikan skripsi ini, sebagai salah satu syarat untuk mendapatkan gelar kesarjanaan. yang dimana jauh dari kata sempurna, namun penulis bangga telah mencapai pada titik ini, yang akhirnya skripsi ini bisa selesai.

Terimakasih juga kepada diri sendiri karena sudah bisa melewati semua ini dengan penuh suka dan duka. penulis selalu menanamkan prinsip apa yang sudah dimulai maka selesaikanlah apapun rintangan-nya. sehingga hal inilah yang membuat penulis memacu dirinya sampai batas maksimal sehingga dapat menyelesaikan skripsi ini.

Skripsi ini penulis persembahkan kepada:

1. Tuhan Yang Maha Esa, atas rahmat dan karunianya yang telah diberikan sehingga penulis dapat menyelesaikan skripsi ini.
2. Kedua Orang tua saya tercinta Bapak Gaosul Alam dan Ibu Rifah yang senantiasa mendoakan anaknya tanpa henti sehingga tidak bosan memberikan semangat, motivasi, pengorbanan, nasihat serta kasih sayang yang tidak pernah henti sampai saat ini.
3. Dosen Pembimbing Bapak Robert Marco yang senantiasa selalu membimbing serta memberikan masukan, nasihat dan saran selama masa penulisan skripsi ini, sehingga penulis dapat menyelesaikan skripsi ini.
4. Sahabat seperjuangan yang memberikan semangat dan juga motivasinya untuk mengerjakan skripsi ini.
5. Semua teman-teman penulis yang berada di Yogyakarta yang selalu menemani ketika penulis sedang tidak baik baik saja.

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Allah SWT, atas nikmat berkah, rahmat dan karunianya sehingga penulis dapat menyelesaikan skripsi dengan judul **“DETEKSI KECURANGAN KEUANGAN DENGAN MENGGUNAKAN METODE RANDOM FOREST DAN XGBOOST PADA DATA YANG TIDAKSEIMBANG”**

Pada kesempatan ini penulis juga ingin menyampaikan berbagai terima kasih yang sebesar-besarnya kepada semua pihak yang sudah memberikan bimbingan, masukan, semangat, nasihat, motivasi dan dorongan untuk menyelesaikan skripsi ini, yaitu kepada:

1. Ibu Prof. Dr. Kusriani., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
2. Bapak Dony Ariyus, S.S., M.Kom selaku Ketua Program Studi Teknik Komputer Universitas Amikom Yogyakarta.
3. Bapak Robert Marco, M. T., Ph.D selaku Dosen Pembimbing yang senantiasa selalu memberikan arahan, saran, motivasi, dan waktunya kepada penulis selama menyusun skripsi ini.
4. Kedua Orang Tua penulis yang senantiasa memberikan semangat, motivasi, nasihat dan dukungan guna kelancaran penyusunan skripsi ini.
5. Semua pihak yang tidak dapat disebutkan satu-persatu yang telah membantu dalam penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih terdapat banyak kekurangan. Oleh karena itu, penulis mengharap kritik dan saran yang membangun guna menyempurnakan skripsi ini. Penulis berharap skripsi ini dapat bermanfaat bagi berbagai pihak.

Yogyakarta, 3 Desember 2025

Penulis

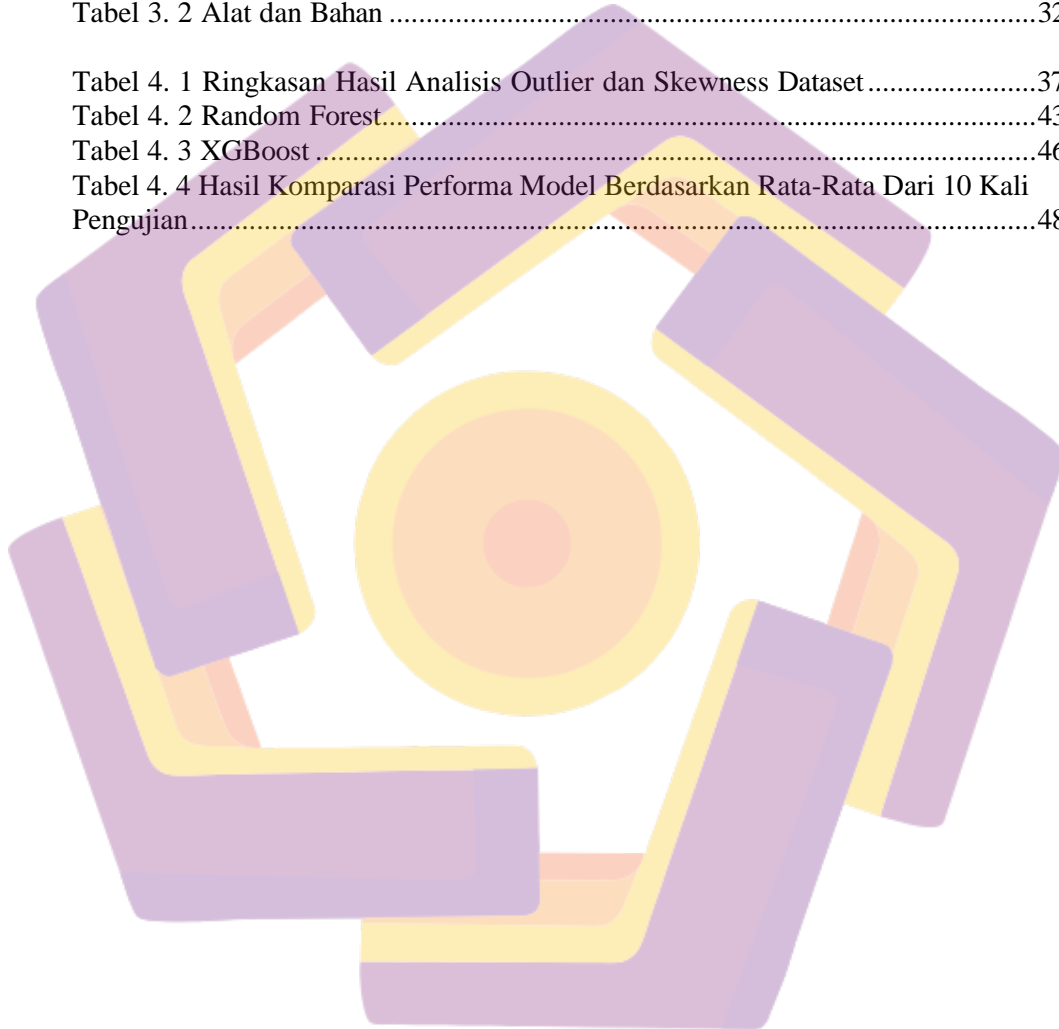
DAFTAR ISI

<u>HALAMAN PERSETUJUAN</u>	ii
<u>HALAMAN PENGESAHAN</u>	iii
<u>HALAMAN PERNYATAAN KEASLIAN SKRIPSI</u>	iv
<u>HALAMAN PERSEMBAHAN</u>	v
<u>KATA PENGANTAR</u>	vi
<u>DAFTAR ISI</u>	vii
<u>DAFTAR TABEL</u>	ix
<u>DAFTAR GAMBAR</u>	x
<u>DAFTAR LAMBANG DAN SINGKATAN</u>	xi
<u>DAFTAR ISTILAH</u>	xii
<u>INTISARI</u>	xiv
<u>ABSTRACT</u>	xv
<u>BAB I PENDAHULUAN</u>	1
<u>1.1 Latar Belakang</u>	1
<u>1.2 Rumusan Masalah</u>	2
<u>1.3 Batasan Masalah</u>	2
<u>1.4 Tujuan Penelitian</u>	3
<u>1.5 Sistematika Penulisan</u>	5
<u>BAB II TINJAUAN PUSTAKA</u>	6
<u>2.1 Studi Literatur</u>	6
<u>2.2 Dasar Teori</u>	11
<u>BAB III METODE PENELITIAN</u>	23
<u>3.1 Objek Penelitian</u>	23

<u>3.2</u>	<u>Alur Penelitian</u>	23
<u>3.2.1</u>	<u>Eksplorasi awal</u>	25
<u>3.2.2</u>	<u>Preprocessing</u>	25
<u>3.2.3</u>	<u>Data Training</u>	27
<u>3.2.4</u>	<u>Model Klasifikasi</u>	27
<u>3.2.5</u>	<u>Evaluasi Model</u>	29
<u>3.3</u>	<u>Alat dan Bahan</u>	30
<u>BAB IV HASIL DAN PEMBAHASAN</u>		33
<u>4.1.</u>	<u>Hasil</u>	33
<u>4.1.1</u>	<u>Eksplorasi Data</u>	33
<u>4.1.2</u>	<u>Preprocessing</u>	36
<u>4.1.3</u>	<u>Outlier and Skewness detection</u>	36
<u>4.1.4</u>	<u>Undersampling Manual Nilai pada Fitur ‘class’</u>	40
<u>4.1.5</u>	<u>Model Klasifikasi</u>	42
<u>4.1.6</u>	<u>Random Forest</u>	43
<u>4.1.7</u>	<u>XGBoost</u>	46
<u>4.2.</u>	<u>Pembahasan</u>	48
<u>BAB V PENUTUP</u>		50
<u>5.1</u>	<u>Kesimpulan</u>	50
<u>5.2</u>	<u>Saran</u>	50
<u>REFERENSI</u>		52
<u>LAMPIRAN A</u>		56

DAFTAR TABEL

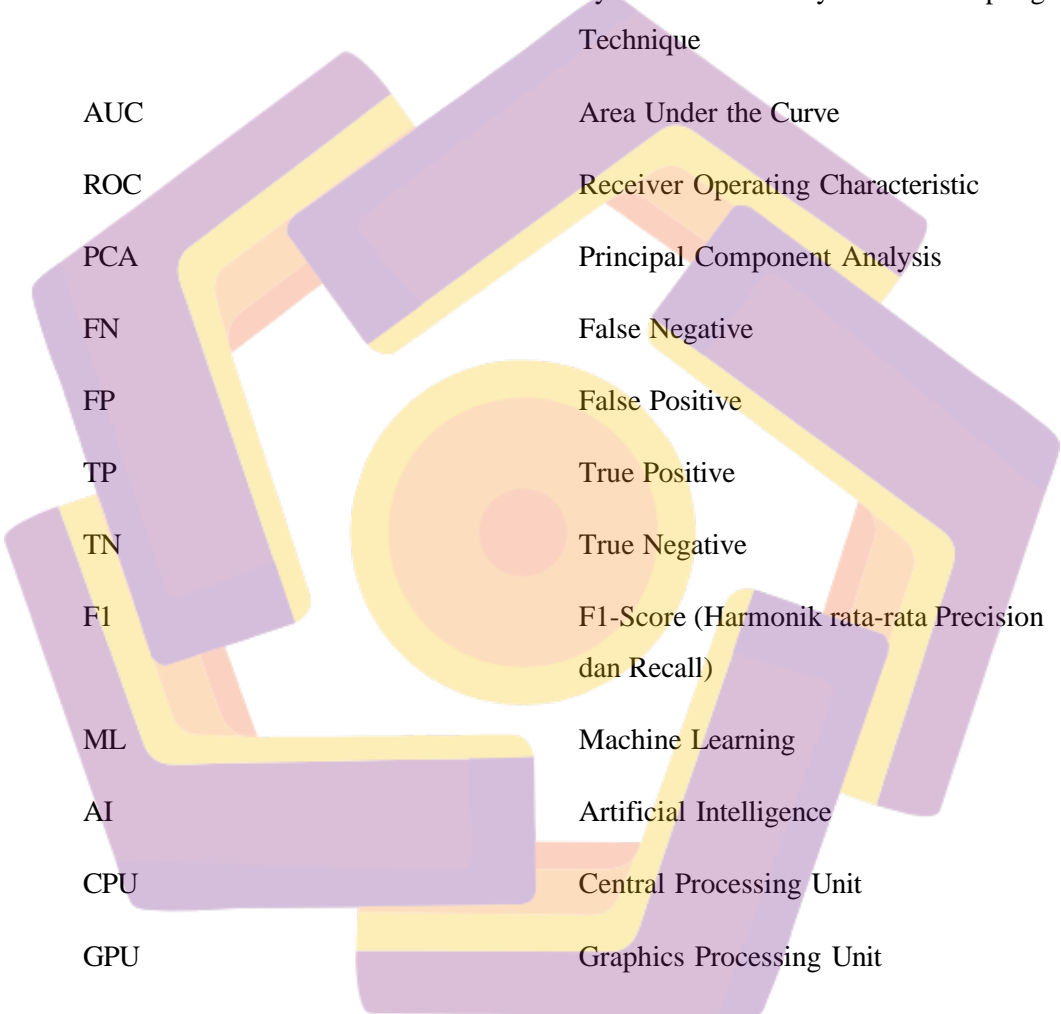
Tabel 2. 1 Keaslian Penelitian	9
Tabel 2. 2 Confusion Matrix.....	20
Tabel 3. 1 Data Penelitian.....	31
Tabel 3. 2 Alat dan Bahan	32
Tabel 4. 1 Ringkasan Hasil Analisis Outlier dan Skewness Dataset.....	37
Tabel 4. 2 Random Forest.....	43
Tabel 4. 3 XGBoost	46
Tabel 4. 4 Hasil Komparasi Performa Model Berdasarkan Rata-Rata Dari 10 Kali Pengujian.....	48



DAFTAR GAMBAR

Gambar 2. 1 Confusion Matrix.....	20
Gambar 3. 1 Flowchart Metodologi Penelitian	24
Gambar 4. 1 Distribution of Transaction Times.....	34
Gambar 4. 2 Tingkat Korelasi Absolut Antar Fitur	35
Gambar 4. 3 Outline Detection V12.....	38
Gambar 4. 4 Outline Detection V14.....	39
Gambar 4. 5 Outline Detection V17.....	39
Gambar 4. 6 Hasil Undersampling pada Fitur “Class”	41
Gambar 4. 7 Distribusi Data Setelah Oversampling SMOTE	42
Gambar 4. 8 Precision-Recall Curve Random Forest.....	44
Gambar 4. 9 Precision-Recall Curve XGBoost	47

DAFTAR LAMBANG DAN SINGKATAN



RF	Random Forest
XGBoost	Extreme Gradient Boosting
SMOTE	Synthetic Minority Over-sampling Technique
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
PCA	Principal Component Analysis
FN	False Negative
FP	False Positive
TP	True Positive
TN	True Negative
F1	F1-Score (Harmonik rata-rata Precision dan Recall)
ML	Machine Learning
AI	Artificial Intelligence
CPU	Central Processing Unit
GPU	Graphics Processing Unit
IQR	Interquartile Range
ENN	Edited Nearest Neighbor

DAFTAR ISTILAH



Bagging	Teknik ensemble yang menggunakan pengambilan sampel bootstrap untuk membuat beberapa subset data pelatihan guna meningkatkan stabilitas model.
Boosting	Metode ensemble yang menggabungkan beberapa model lemah secara berurutan untuk memperbaiki kesalahan prediksi sebelumnya.
Oversampling	Proses menambah sampel kelas minoritas agar distribusi data lebih seimbang.
Underfitting	Kondisi ketika model terlalu sederhana sehingga tidak mampu menangkap pola data dengan baik.
Overfitting	Kondisi ketika model terlalu menyesuaikan diri dengan data latih sehingga tidak bekerja baik pada data baru.
Cross-validation	Teknik evaluasi model dengan membagi dataset menjadi beberapa bagian untuk melatih dan menguji model secara bergantian.
Precision	Rasio prediksi positif yang benar terhadap semua prediksi positif.

Recall

Rasio jumlah kasus positif yang berhasil dideteksi model terhadap semua kasus positif sebenarnya.

F1-Score

Nilai harmonik antara Precision dan Recall, digunakan untuk mengukur keseimbangan kinerja model.

Imbalanced Data

Kondisi dataset di mana jumlah data antar kelas berbeda jauh, misalnya kasus fraud yang jauh lebih sedikit dibanding non-fraud.

Outlier

Data dengan nilai ekstrem yang menyimpang dari distribusi umum dan dapat memengaruhi hasil analisis.

Feature Engineering

Proses pembuatan atau transformasi fitur untuk meningkatkan kemampuan model mempelajari pola data.

Normalization

Proses penskalaan nilai data agar memiliki rentang yang seragam untuk mempercepat proses pembelajaran model.

INTISARI

Penelitian ini bertujuan mendeteksi fraud dengan teknik ensemble machine learning, khususnya Random Forest (RF) dan Extreme Gradient Boosting (XGBoost), diterapkan pada dataset Credit Card Fraud Detection Kaggle (284.807 sampel, 0,17% fraud). Proses metodologi mencakup preprocessing melalui normalisasi dan Synthetic Minority Over-sampling Technique (SMOTE) untuk menangani ketidakseimbangan, diikuti pelatihan model menggunakan Scikit-learn serta XGBoost di Python. RF efektif mencegah overfitting lewat bagging, sementara XGBoost unggul dalam optimasi gradient untuk data skala besar. Evaluasi dilakukan dengan metrik Area Under the Curve (AUC-ROC), precision, recall, dan F1-score melalui cross-validation. Hasil menunjukkan XGBoost lebih superior (AUC-ROC 0,98; precision 0,95; recall 0,92; F1-score 0,93) dibanding RF (0,96; 0,92; 0,89; 0,90), berkat ketangguhannya terhadap noise. Penelitian menyimpulkan XGBoost sebagai metode optimal untuk data imbalanced, dengan saran integrasi SMOTE guna tingkatkan recall hingga 15%. Kontribusi teoritis memperkaya literatur machine learning di keamanan siber (Al-Hashedi & Magalingam, 2021), sementara manfaat praktisnya mengurangi kerugian bagi bank dan fintech, serta mendukung pengembangan model real-time.

Kata kunci: *Deteksi Fraud, Ensemble Machine Learning, Random Forest, XGBoost, Data Imbalanced.*

ABSTRACT

This study aims to detect fraud using ensemble machine learning techniques, specifically Random Forest (RF) and Extreme Gradient Boosting (XGBoost), applied to the Credit Card Fraud Detection Kaggle dataset (284,807 samples, 0.17% fraud). The methodological process includes preprocessing through normalization and Synthetic Minority Over-sampling Technique (SMOTE) to address imbalance, followed by model training using Scikit-learn and XGBoost in Python. RF effectively prevents overfitting through bagging, while XGBoost excels in gradient optimization for large-scale data. The evaluation was conducted using the Area Under the Curve (AUC-ROC), precision, recall, and F1-score metrics through cross-validation. The results show that XGBoost is superior (AUC-ROC 0.98; precision 0.95; recall 0.92; F1-score 0.93) compared to RF (0.96; 0.92; 0.89; 0.90), thanks to its robustness to noise. The study concluded that XGBoost is the optimal method for imbalanced data, with the recommendation to integrate SMOTE to increase recall by up to 15%. The theoretical contribution enriches the machine learning literature on cybersecurity (Al-Hashedi & Magalingam, 2021)[1], while the practical benefits reduce losses for banks and fintech companies and support the development of real-time models.

Keyword: *Fraud Detection, Ensemble Machine Learning, Random Forest, XGBoost, Imbalanced Data.*