

**PERBANDINGAN EFEKTIVITAS ALGORITMA *RANDOM FOREST*
DAN *NAIVE BAYES* DALAM DETEKSI PENIPUAN PADA
*E-COMMERCE***

SKRIPSI

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



disusun oleh
RISKA SEFIANA PUTRI
22.11.5055

Kepada
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

**PERBANDINGAN EFEKTIVITAS ALGORITMA *RANDOM*
FOREST DAN *NAIVE BAYES* DALAM DETEKSI PENIPUAN
*E-COMMERCE***

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
RISKA SEFIANA PUTRI
22.11.5055

Kepada
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN EFEKTIVITAS ALGORITMA *RANDOM FOREST* DAN
NAIVE BAYES DALAM DETEKSI PENIPUAN PADA *E-COMMERCE***

yang disusun dan diajukan oleh

Riska Sefiana Putri

22.11.5055

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Januari 2026

Dosen Pembimbing,



Anna Baita, S. Kom., M.Kom

NIK. 190302290

HALAMAN PENGESAHAN
SKRIPSI

**PERBANDINGAN EFEKTIVITAS ALGORITMA *RANDOM FOREST* DAN
NAIVE BAYES DALAM DETEKSI PENIPUAN PADA *E-COMMERCE***

yang disusun dan diajukan oleh

Riska Sefiana Putri

22.11.5055

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega P. D., S. Kom., M. Kom.,
NIK. 190302185



Nafiatun Sholihah, S. Kom., M. Cs.,
NIK. 190302524



Anna Balta, S. Kom., M. Kom.,
NIK. 190302290



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Riska Sefiana Putri
NIM : 22.11.5055

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan Efektivitas Algoritma *Random Forest* dan *Naive Bayes* dalam Deteksi Penipuan Pada *E-Commerce*

Dosen Pembimbing : Anna Baita, 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, 23 Januari 2026

Yang Menyatakan,



Riska Sefiana Putri

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT atas segala rahmat dan kemudahan yang diberikan, skripsi ini saya persembahkan kepada:

1. Kedua orang tua tercinta, yang senantiasa memberikan doa, kasih sayang, dukungan, serta pengorbanan tanpa batas dalam setiap langkah kehidupan penulis. Segala pencapaian ini tidak terlepas dari bimbingan, kepercayaan, dan ketulusan yang selalu diberikan. Terima kasih atas kesabaran, motivasi, dan semangat yang terus menguatkan penulis dalam menyelesaikan pendidikan hingga tahap akhir ini.
2. Kakak dan adik tercinta, yang senantiasa memberikan motivasi, dukungan, dan semangat dalam setiap proses yang penulis jalani. Kehadiran dan perhatian kalian menjadi penguat bagi penulis untuk tetap konsisten hingga menyelesaikan skripsi ini.
3. Keluarga besar penulis, atas doa, dukungan, dan perhatian yang diberikan selama masa perkuliahan hingga terselesaikannya skripsi ini.
4. Sahabat dan teman-teman seperjuangan, yang telah memberikan kebersamaan, dukungan, serta semangat selama proses perkuliahan dan penyusunan skripsi. Diskusi, kerja sama, serta motivasi yang diberikan menjadi bagian dari perjalanan yang penuh pembelajaran dan pengalaman berharga.
5. Diri penulis sendiri, Riska Sefiana Putri, sebagai bentuk penghargaan atas keberanian untuk bertahan, berproses, dan terus belajar selama masa perkuliahan. Penyusunan skripsi ini bukan hanya pemenuhan kewajiban akademik, tetapi juga wujud ketekunan dan tanggung jawab dalam mencapai tujuan.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Perbandingan Efektivitas Algoritma Random Forest dan Naive Bayes dalam Deteksi Penipuan pada E-Commerce" dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Universitas AMIKOM Yogyakarta.

Dalam proses penyusunan skripsi ini, penulis memperoleh banyak bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan terima kasih kepada:

1. Ibu Anna Baita, S.Kom., M.Kom., selaku Dosen Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi dengan penuh kesabaran selama proses penyusunan skripsi ini.
2. Ibu Windha Mega P. D., S.Kom., M.Kom., selaku Dosen Penguji I yang telah memberikan kritik dan saran demi penyempurnaan skripsi ini.
3. Ibu Nafi'atun Sholihah, S.Kom., M.Cs., selaku Dosen Penguji II yang telah memberikan masukan dan arahan yang membangun.
4. Seluruh Dosen Program Studi Informatika Universitas AMIKOM Yogyakarta yang telah memberikan ilmu dan wawasan selama masa perkuliahan.
5. Kedua orang tua dan keluarga penulis yang senantiasa memberikan doa, dukungan, dan semangat dalam menyelesaikan studi.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan pengembangan ilmu pengetahuan.

Yogyakarta, 23 Januari 2026

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
ABSTRACT	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	12

2.2.1 <i>E-Commerce</i>	12
2.2.2 Deteksi Penipuan (<i>Fraud Detection</i>)	13
2.2.3 <i>Machine Learning</i>	13
2.2.4 <i>Preprocessing Data</i>	15
2.2.5 <i>Exploratory Data Analysis</i>	16
2.2.6 <i>Feature Engineering</i>	17
2.2.7 <i>One-hot Encoding</i>	18
2.2.8 <i>Feature selection</i>	18
2.2.8 Normalisasi	23
2.2.8 <i>Imbalanced Data</i>	24
2.2.9 <i>Undersampling</i>	24
2.2.10 <i>Random Under Sampling (RUS)</i>	25
2.2.9 <i>Random Forest</i>	26
2.2.10 <i>Naïve Bayes</i>	27
2.2.11 Evaluasi Model	31
BAB III METODE PENELITIAN	36
3.1 Objek Penelitian	36
3.3 Alat dan Bahan	36
3.3.1 Data Penelitian	36
3.3.2 Alat/Instrumen	37
3.2 Alur Penelitian	39
BAB IV HASIL DAN PEMBAHASAN	45
4.1 Data Acquisition	45
4.2 <i>Exploratory Data Analysis (EDA)</i>	46
4.2 <i>Preprocessing</i>	52

4.3 <i>Feature Engineering</i>	55
4.4 <i>Feature selection</i>	57
4.5 Data Split	60
4.6 Ketidakseimbangan Kelas	60
4.7 Normalisasi	61
4.7 Modeling	62
4.7.1 <i>Baseline Modeling</i>	63
4.7.2 <i>Hyperparameter Tuning</i>	65
4.9 Analisis Komputasi	69
4.10 Evaluasi Model	71
4.10.1 Hasil Studi Komparasi	71
4.11 <i>Deployment</i>	73
BAB V PENUTUP	75
5.1 Kesimpulan	75
5.2 Saran	76
REFERENSI	77

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	8
Tabel 4. 1 Deskripsi Fitur Dataset Transaksi <i>E-Commerce</i>	45
Tabel 4. 2 Distribusi <i>Outlier</i> Fitur Numerik Berdasarkan Kelas	54
Tabel 4. 3 Hasil <i>Engineering</i>	56
Tabel 4. 4 Keseluruhan Fitur	58
Tabel 4. 5 Hasil <i>Feature selection</i>	59
Tabel 4. 6 Hasil Data Split	60
Tabel 4. 7 Distribusi Kelas Sebelum dan Sesudah <i>Random Under Sampling (RUS)</i>	60
Tabel 4. 8 Hasil <i>Baseline</i> pada Kedua Algoritma	63
Tabel 4. 9 <i>Search Space</i> pada <i>Hyperparameter Tuning</i>	65
Tabel 4. 10 Hasil Best Parameter	66
Tabel 4. 11 Hasil <i>Hyperparameter Tuning</i> kedua Algoritma	67
Tabel 4. 12 Hasil Analisis Komputasi	70
Tabel 4. 13 Hasil Studi Komparasi	72

DAFTAR GAMBAR

Gambar 2. 1 <i>Confusion Matrik</i>	32
Gambar 3. 1 Alur Penelitian A	39
Gambar 3. 2 Alur Penelitian B	40
Gambar 4. 1 Distribusi Kelas <i>Fraud VS Non-Fraud</i>	47
Gambar 4. 2 Distribusi <i>user_id</i>	48
Gambar 4. 3 Distribusi <i>purchase_value</i>	49
Gambar 4. 4 Distribusi <i>age</i>	50
Gambar 4. 5 Distribusi <i>ip_address</i>	51
Gambar 4. 6 <i>Correlation Heatmap</i> Fitur Numerik.	52
Gambar 4. 7 Gambar Kurva ROC CV 3 dan CV 5	65
Gambar 4. 8 Gambar Kurva ROC pada <i>Tuning</i> menggunakan CV 3 dan CV 5	69
Gambar 4. 9 Hasil <i>Deployment</i>	74

DAFTAR LAMBANG DAN SINGKATAN

K	jumlah kelas,
n_k	jumlah data pada kelas ke- k .
$\bar{x}_k$	rata-rata fitur pada kelas ke- k .
$\bar{x}$	rata-rata total fitur.
N	jumlah seluruh data.
h	Hipotesis data dengan suatu <i>class</i> tertentu.
D	Data yang belum memiliki <i>class</i> .
$P(h)$	Probabilitas hipotesis (<i>prior probability</i>).
$P(D)$	Probabilitas D .
C_k	kelas komplement dari kelas C_k .
N_{ij}	jumlah kemunculan fitur ke- i pada kelas j .
α	parameter smoothing
m	jumlah fitur.
ANOVA	Analysis of Variance
AUC	Area Under Curve
CNB	Complement Naive Bayes
CV	Cross Validation
EDA	Exploratory Data Analysis
F1-score	Harmonic Mean Precision dan Recall
ML	Machine Learning
NB	Naive Bayes
RF	Random Forest
ROC	Receiver Operating Characteristic
RUS	Random Under Sampling
TPR	True Positive Rate
FPR	False Positive Rate
FNR	False Negative Rate.

DAFTAR ISTILAH

Accuracy	persentase prediksi yang benar dari seluruh data.
Data cleaning	proses pembersihan data dari duplikasi dan nilai kosong.
Encoding	transformasi data kategorikal menjadi numerik.
EDA	analisis awal untuk memahami karakteristik data.
E-Commerce	sistem transaksi jual beli secara digital.
False Negative Rate	fraud yang salah diklasifikasikan sebagai non-fraud.
False Positive Rate	non-fraud yang salah diklasifikasikan sebagai fraud.
Feature Engineering	pembentukan fitur baru dari data mentah.
Feature Selection	pembentukan fitur yang paling relevan untuk pemodelan.
Hyperparameter Tuning	optimasi parameter
Machine learning	metode pembelajaran mesin dari data.
Min-Max Scaling	normalisasi data ke rentang 0-1.
Naive Bayes	Algoritma klasifikasi berbasis probabilitas.
Precision	ketepatan prediksi fraud.
Random Forest	Algoritma klasifikasi berbasis ensemble pohon keputusan.
RUS	penyeimbangan data dengan mengurangi kelas mayoritas.
Recall	Kemampuan model mendeteksi transaksi fraud.
ROC-AUC	ukuran performa model dalam membedakan kelas.

INTISARI

Pertumbuhan *e-commerce* telah meningkatkan kemudahan transaksi digital, namun di sisi lain juga meningkatkan risiko penipuan (*fraud*) seperti transaksi fiktif, pencurian identitas, dan penyalahgunaan metode pembayaran. Aktivitas penipuan ini menimbulkan kerugian finansial yang signifikan serta menurunkan tingkat kepercayaan pengguna terhadap platform *e-commerce*. Seiring meningkatnya volume dan kompleksitas data transaksi, metode deteksi *fraud* secara manual menjadi tidak efektif, sehingga diperlukan pendekatan berbasis machine learning yang mampu mendeteksi transaksi penipuan secara tepat dan efisien, khususnya pada kondisi data yang tidak seimbang.

Penelitian ini membandingkan kinerja algoritma *Random Forest* dan *Naïve Bayes* dalam mendeteksi transaksi penipuan pada data *e-commerce*. Dataset diproses melalui tahapan pembersihan data, transformasi variabel kategorikal menggunakan *encoding*, normalisasi fitur numerik, serta penanganan ketidakseimbangan kelas menggunakan metode *Random Under Sampling (RUS)*. Evaluasi model dilakukan menggunakan skema *cross-validation* dengan fokus pada metrik deteksi kelas *fraud*.

Hasil penelitian menunjukkan bahwa *Random Forest* memiliki kemampuan deteksi *fraud* yang lebih baik, ditunjukkan oleh nilai *recall* rata-rata sebesar 0,7191, yang berarti model mampu mengidentifikasi sekitar 71,91% transaksi penipuan, serta menghasilkan *False Negative Rate (FNR)* yang lebih rendah, sehingga risiko transaksi *fraud* yang tidak terdeteksi dapat diminimalkan. Sebaliknya, *Naïve Bayes* menghasilkan nilai *recall* sebesar 0,5520, namun menunjukkan efisiensi komputasi yang jauh lebih tinggi, dengan waktu prediksi rata-rata 0,13 detik dan ukuran model yang sangat kecil (<0,01 MB). Temuan ini menunjukkan adanya *trade-off* antara kemampuan deteksi *fraud* dan efisiensi komputasi dalam pemilihan algoritma untuk sistem *e-commerce*.

Kata kunci: *Random Forest*, *Naive Bayes*, Deteksi Penipuan, *E-Commerce*, *Machine Learning*.

ABSTRACT

The rapid growth of e-commerce has increased the convenience of digital transactions, but it has also led to a higher risk of fraudulent activities, such as fictitious transactions, identity theft, and misuse of payment methods. These fraudulent activities cause significant financial losses and reduce user trust in e-commerce platforms. As transaction data continue to grow in volume and complexity, manual fraud detection methods become ineffective, making machine learning-based approaches essential, particularly for imbalanced datasets.

This study compares the performance of Random Forest and Naïve Bayes algorithms for detecting fraudulent transactions in e-commerce data. The dataset was processed through several stages, including data cleaning, categorical feature encoding, numerical feature normalization, and class imbalance handling using Random Under Sampling (RUS). Model evaluation was conducted using a cross-validation scheme, with a primary focus on metrics relevant to fraud detection.

The results show that Random Forest demonstrates superior fraud detection capability, achieving an average recall of 0.7191, indicating that approximately 71.91% of fraudulent transactions were correctly identified, along with a lower False Negative Rate (FNR). In contrast, Naïve Bayes obtained a recall of 0.5520, but exhibited significantly higher computational efficiency, with an average prediction time of 0.13 seconds and a very small model size (<0.01 MB). These findings highlight a clear trade-off between fraud detection effectiveness and computational efficiency in selecting machine learning algorithms for e-commerce fraud detection systems.

Keyword: Random Forest, Naïve Bayes, Fraud Detection, E-Commerce, Machine Learning.