PENERAPAN ALGORITMA RANDOM FOREST UNTUK DETEKSI ANOMALI SERANGAN PHISHING PADA JARINGAN KOMPUTER

SKRIPSI

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

disusun oleh

BEN RAFI KAHMAS

21.11.3873

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

PENERAPAN ALGORITMA RANDOM FOREST UNTUK DETEKSI ANOMALI SERANGAN PHISHING PADA JARINGAN KOMPUTER

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika

disusun oleh

BEN RAFI KAHMAS

21.11.3873

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

HALAMAN PERSETUJUAN

SKRIPSI

PENERAPAN ALGORITMA RANDOM FOREST UNTUK DETEKSI ANOMALI SERANGAN PHISHING PADA JARINGAN KOMPUTER

yang disusun dan diajukan oleh

Ben Rafi Kahmas

21.11.3873

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Agustus 2025

Dosen Pembimbing,

Andika Agus Slameto, M.Kom.
NIK. 190302109

HALAMAN PENGESAHAN

SKRIPSI

PENERAPAN ALGORITMA RANDOM FOREST UNTUK DETEKSI ANOMALI SERANGAN PHISHING PADA JARINGAN KOMPUTER

yang disusun dan diajukan oleh

Ben Rafi Kalmas

21.11.3873

Telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Agustus 2025

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Bayu Setiaji, M.Kom. **NIK. 190302216**	
Bambang Pilu Hartato, S.Kom., M.Eng. **NIK. 190302707**	
Andika Agus Slameto, M.Kom. **NIK. 190302169**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ben Rafi Kahmas
NIM : 21.11.3873

Menyatakan bahwa Skripsi dengan judul berikut:

Penerapan Algoritma Random Forest Untuk Deteksi Anomali Serangan Phishing Pada Jaringan Komputer

Dosen Pembimbing : Andika Agus Slameto, 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, 26 Agustus 2025

Yang Menyatakan,

Ben Rafi Kahmas

HALAMAN PERSEMBAHAN

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, karunia, dan ridho-Nya. Berkat kesehatan, kekuatan, serta kelancaran yang diberikan, penulis akhirnya dapat menyelesaikan penyusunan skripsi ini dengan baik. Dengan penuh rasa hormat dan terima kasih, penulis menyampaikan apresiasi yang sebesar-besarnya kepada pihak-pihak yang telah memberikan dukungan selama proses penyusunan skripsi ini:

1. Kedua orang tua tercinta, Alm. Bowo Cipto Hadi dan Ibu Kusmiyati, serta kakak-kakak saya, Gun Bondan Wiratama dan Habib Rido Pangestu, atas segala doa, dorongan, dan semangat yang selalu menyertai penulis dalam setiap langkah perjuangan.
2. Dosen pembimbing, Bapak Andika Agus Slameto, M.Kom., yang dengan sabar dan penuh dedikasi telah membimbing penulis sejak awal hingga tuntasnya skripsi ini.
3. Rekan-rekan kelas Informatika 01 Universitas Amikom Yogyakarta angkatan 2021, yang telah banyak memberikan bantuan, motivasi, dan kebersamaan selama proses penyusunan skripsi ini.

KATA PENGANTAR

Segala puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, karunia, dan ridho-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Penerapan Algoritma Random Forest Untuk Deteksi Anomali Serangan Phishing Pada Jaringan Komputer". Shalawat serta salam juga senantiasa tercurah kepada junjungan Nabi Muhammad shallallahu 'alaihi wasallam, beserta keluarga, para sahabat, dan seluruh pengikut setia beliau.

Skripsi ini disusun sebagai salah satu syarat untuk meraih gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer. Dalam proses penyusunan skripsi ini, penulis mendapatkan banyak dukungan, bimbingan, masukan, dan bantuan dari berbagai pihak, baik berupa saran, kritik yang membangun, semangat, maupun bantuan teknis lainnya. Oleh karena itu, dengan penuh rasa hormat dan penghargaan, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Kedua orang tua tercinta, Alm. Bowo Cipto Hadi dan Ibu Kusmiyati, serta kakak-kakak saya, Gun Bondan Wiratama dan Habib Rido Pangestu, atas segala doa, dorongan, dan semangat yang selalu menyertai penulis dalam setiap langkah perjuangan.
2. Bapak Prof. Dr. M. Suyanto, M.M, selaku Rektor Universitas AMIKOM Yogyakarta.
3. Bapak Prof. Dr. Kusrini, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta beserta seluruh jajarannya.
4. Ibu Eli Pujiastuti, M.Kom, selaku Ketua Program Studi Informatika Universitas AMIKOM Yogyakarta beserta seluruh jajarannya.
5. Bapak Andika Agus Slameto, M.Kom., selaku dosen pembimbing penulis yang telah memberikan kritik, saran, waktu, motivasi dan bimbingan dalam skripsi ini sehingga skripsi ini dapat selesai dengan baik dan tepat waktu.
6. Bapak -, dan ibu -., selaku dosen penguji yang telah memberikan saran dan masukan atas penyelesaian skripsi ini.

7. Bapak dan Ibu Dosen Program Studi Informatika Universitas AMIKOM Yogyakarta yang telah mengajari penulis dari semester awal hingga akhir sehingga penulis memperoleh banyak sekali ilmu yang bermanfaat dan pengalaman yang sangat berharga.
8. Rekan-rekan Informatika 01 Universitas Amikom Yogyakarta angkatan 2021, yang telah banyak memberikan bantuan, motivasi, dan kebersamaan selama proses penyusunan skripsi ini.

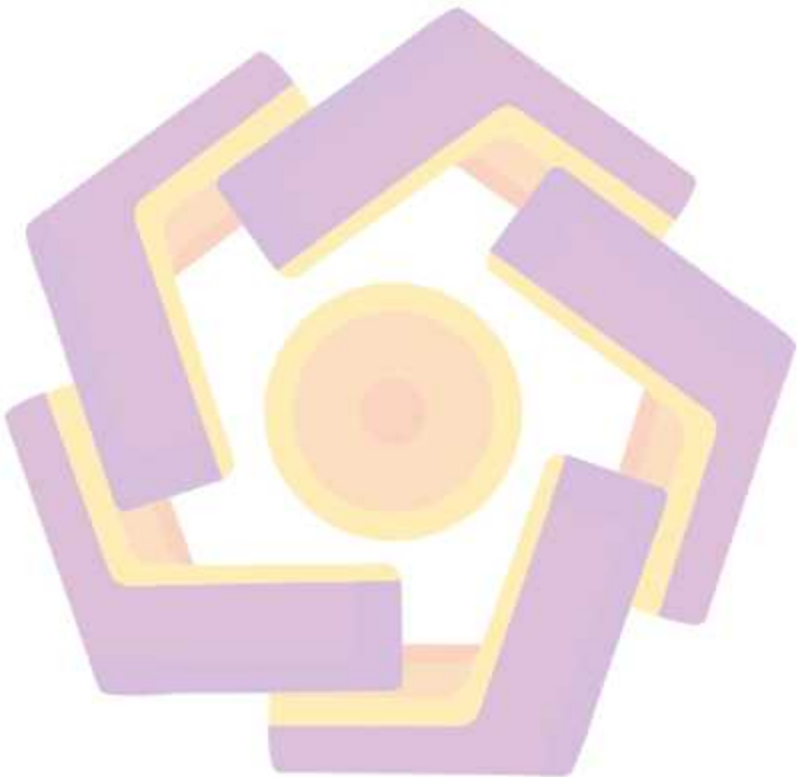
Yogyakarta, 26 Agustus 2025
Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

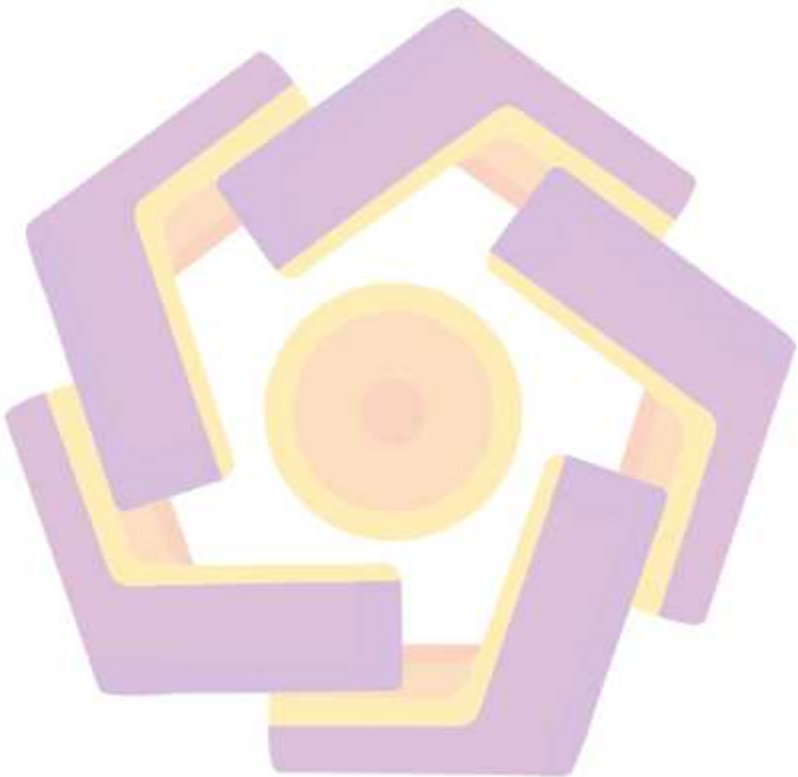
DAFTAR LAMPIRAN



DAFTAR LAMBANG DAN SINGKATAN

CNN	Convolutional Neural Network
CT	Classification Tree
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DNS	Domain Name System
DT	Decision Tree
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IP	Internet Protocol
KNN	K-Nearest Neighbor
LR	Logistic Regression
PCA	Principal Component Analysis
RF	Random Forest
SVM	Support Vector Machine
TCP	Transmission Control Protocol
TLS	Transport Layer Security
URL	Uniform Resource Locator
XGBoost	Extreme Gradient Boosting
$\hat{y}$	Prediksi akhir
n	Jumlah pohon atau kelas
$T_i(x)$	Prediksi dari pohon ke-i untuk input x
I_G	Gini index
P_i	Proporsi sampel dari kelas i dalam node
I_H	Entropy

$\sum$ Sigma



DAFTAR ISTILAH

Accuracy	Ukuran seberapa banyak prediksi model yang benar dibandingkan total prediksi.
Artificial Intelligent	Kecerdasan buatan yang dirancang untuk meniru kemampuan berpikir manusia.
Blacklisting	Metode keamanan dengan memblokir domain, IP, atau URL yang dianggap berbahaya.
Boxplot	Visualisasi statistik untuk menunjukkan distribusi data, median, dan outlier.
Confusion Matrix	Matriks evaluasi yang menunjukkan jumlah prediksi benar dan salah untuk tiap kelas.
Dataset	Kumpulan data yang digunakan untuk pelatihan atau pengujian dalam machine learning.
Domain	Alamat unik dari suatu situs web (misal: example.com).
Elemen Redirect	Elemen pada halaman web yang mengarahkan pengguna ke halaman lain secara otomatis.
Entropy	Ukuran ketidakpastian dalam pembentukan pohon keputusan.
F-measure	Rata-rata harmonik dari presisi dan recall, digunakan untuk evaluasi model.
Firewall	Sistem keamanan jaringan yang memfilter lalu lintas berdasarkan aturan tertentu.
Gini Index	Metode untuk mengukur kemurnian node dalam pohon keputusan.
Google Colab	Platform cloud berbasis Jupyter Notebook untuk menjalankan kode Python.
Heatmap Correlation	Visualisasi hubungan antar variabel dalam bentuk peta warna.

Histogram	Grafik batang untuk menampilkan distribusi frekuensi data numerik.
Hyperparameter	Parameter yang diatur sebelum pelatihan model dan memengaruhi proses pembelajaran.
Imbalance Data	Kondisi saat jumlah sampel antar kelas dalam dataset tidak seimbang.
Internet	Jaringan global yang menghubungkan komputer dan perangkat di seluruh dunia.
IP Address	Alamat numerik yang mengidentifikasi perangkat dalam jaringan.
Kaggle	Platform berbagi dataset dan kompetisi machine learning.
Log-transform	Teknik transformasi data untuk memperbaiki distribusi yang tidak normal.
Machine Learning	Cabang AI yang memungkinkan komputer belajar dari data tanpa diprogram secara eksplisit.
Noise	Data yang tidak relevan atau mengganggu yang dapat memengaruhi hasil analisis.
Outlier	Nilai yang menyimpang jauh dari pola data lainnya.
Overfitting	Kondisi ketika model terlalu menyesuaikan data latih dan gagal memprediksi data baru.
Parameter	Nilai internal dalam model machine learning yang dipelajari selama pelatihan.
Payload Application	Konten atau bagian dari serangan phishing yang dirancang untuk mencuri informasi.
Wrapper	Metode seleksi fitur yang mengevaluasi kombinasi fitur berdasarkan kinerja model.
PhisTank	Basis data atau layanan online yang menyediakan informasi tentang situs phishing.

Phishing	Upaya penipuan untuk mencuri informasi pribadi dengan menyamar sebagai entitas tepercaya.
Preprocessing Data	Tahap awal pengolahan data untuk mempersiapkannya sebelum digunakan oleh model.
Precison	Proporsi prediksi positif yang benar terhadap seluruh prediksi positif.
Proxy	Server perantara antara pengguna dan internet, sering digunakan untuk keamanan dan filter.
Recall	Proporsi data positif yang berhasil dikenali dengan benar oleh model.
Scatter Plot	Grafik sebaran titik untuk melihat hubungan antara dua variabel numerik.
Server	Komputer yang menyediakan layanan atau sumber daya untuk perangkat lain di jaringan.
Signature Based Detection	Deteksi ancaman berdasarkan pola atau tanda tangan digital yang sudah dikenali.
Spoofing	Teknik pemalsuan identitas untuk menyamar sebagai sumber tepercaya.
Standard Deviasi	Ukuran sebaran atau variasi dari sekumpulan data.
Teknik Oversampling	Menambahkan data sintetis pada kelas minoritas untuk mengatasi data tidak seimbang.
Teknik Undersampling	Mengurangi jumlah data dari kelas mayoritas untuk menyeimbangkan dataset.
Testing Data	Data yang digunakan untuk menguji performa model setelah pelatihan selesai.
Traffic Jaringan	Arus data yang mengalir melalui jaringan komputer.
Training Data	Data yang digunakan untuk melatih model machine learning.

Tuning Model	Proses mengatur parameter atau hyperparameter agar model bekerja optimal.
Typosquatting	Teknik phishing dengan membuat domain yang mirip domain asli tapi dengan kesalahan ejaan.
UCI	Singkatan dari University of California, Irvine – penyedia kumpulan dataset populer.
Zero Day Attack	Serangan yang memanfaatkan celah keamanan yang belum diketahui atau ditambal.

INTISARI

Serangan *phishing* merupakan salah satu ancaman keamanan siber yang paling umum dan berbahaya, karena memanfaatkan teknik manipulasi untuk mencuri informasi sensitif pengguna. Dalam penelitian ini, diusulkan penerapan algoritma *Random Forest* untuk mendeteksi anomali serangan *phishing* pada jaringan komputer. Algoritma *Random Forest* dipilih karena kemampuannya dalam melakukan klasifikasi dengan akurasi tinggi, serta keunggulannya dalam menangani data yang kompleks dan beragam. *Dataset* yang digunakan terdiri dari data lalu lintas jaringan yang mencakup aktivitas normal dan aktivitas yang terindikasi *phishing*. Data kemudian diproses melalui tahap *preprocessing*, pemilihan fitur, dan pelatihan model menggunakan *Random Forest*. Hasil pengujian menunjukkan bahwa model ini mampu mencapai akurasi sebesar 98%, dengan nilai *precision* 98%, *recall* 98%, dan *F1-score* 98%. Penelitian ini juga menunjukkan bahwa fitur-fitur URL seperti persentase tautan eksternal yang mengarahkan kembali ke *domain* awal, ketidaksesuaian nama *domain* yang sering terjadi, jumlah tanda hubung (-) dalam URL, serta indikasi pengiriman data melalui email merupakan fitur yang relevan dan efektif dalam membedakan URL *phishing* dan *non-phishing*. Temuan ini membuktikan bahwa *Random Forest* dapat menjadi metode yang efektif untuk mengidentifikasi serangan *phishing* berdasarkan URL.

Kata kunci: Random Forest, Deteksi Anomali, Phishing, Jaringan Komputer, Pembelajaran Mesin

ABSTRACT

Phishing attacks are among the most common and dangerous cybersecurity threats, as they exploit manipulation techniques to steal sensitive user information. This study proposes the application of the Random Forest algorithm to detect phishing attack anomalies in computer networks. Random Forest was selected for its ability to perform high-accuracy classification and its strength in handling complex and diverse data. The dataset used in this study consists of network traffic data that includes both normal activities and activities identified as phishing. The data underwent preprocessing, feature selection, and model training using Random Forest. The experimental results show that the model achieved 98% accuracy, with precision, recall, and F1-score all reaching 98%. This study also reveals that URL features such as the percentage of external links redirecting back to the original domain, frequent domain name mismatches, the number of hyphens (-) in the URL, and the presence of data submission via email are relevant and effective in distinguishing phishing from non-phishing URLs. These findings confirm that Random Forest can serve as an effective method for identifying phishing attacks based on URL characteristics.

Keyword: *Random Forest, Anomaly Detection, Phishing, Computer Networks, Machine Learning*