DETEKSI WEBSITE PHISHING BERBASIS URL MENGGUNAKAN PENDEKATAN MACHINE LEARNING

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Teknik Komputer

disusun oleh

NUR FAUZI WIBOWO

21.83.0650

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

DETEKSI WEBSITE PHISHING BERBASIS URL MENGGUNAKAN PENDEKATAN MACHINE LEARNING

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Teknik Komputer

disusun oleh

NUR FAUZI WIBOWO

21.83.0650

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

HALAMAN PERSETUJUAN

SKRIPSI

DETEKSI WEBSITE PHISHING BERBASIS URL MENGGUNAKAN PENDEKATAN MACHINE LEARNING

yang disusun dan diajukan oleh

Nur Fauzi Wibowo

21.83.0650

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Juni 2025

Dosen Pembimbing,

Dr. Dony Ariyus, S.S., M.Kom.
NIK. 190302128

HALAMAN PENGESAHAN

SKRIPSI

DETEKSI WEBSITE PHISHING BERBASIS URL MENGGUNAKAN PENDEKATAN MACHINE LEARNING

yang disusun dan diajukan oleh

Nur Fauzi Wibowo

21.83.0650

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Juni 2025

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Dr. Ferry Wahyu Wibowo, S.Si., M.Cs. **NIK. 190302235**	
Muhammad Koprawi, S.Kom., M.Eng. **NIK. 190302454**	
Dr. Dony Ariyus, S.S., M.Kom. **NIK. 190302128**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Juni 2025

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nur Fauzi Wibowo
NIM : 21.83.0650

Menyatakan bahwa Skripsi dengan judul berikut:

DETEKSI WEBSITE PHISHING BERBASIS URL MENGGUNAKAN PENDEKATAN MACHINE LEARNING

Dosen Pembimbing : Dr. Dony Ariyus, S.S., 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 Juni 2025

Yang Menyatakan,

Nur Fauzi Wibowo

HALAMAN PERSEMBAHAN

Segala puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa atas rahmat, karunia, dan kekuatan-Nya yang tiada henti, sehingga saya dapat menyelesaikan tugas akhir ini dengan penuh perjuangan dan keikhlasan. Dengan kerendahan hati dan rasa syukur yang dalam, saya persembahkan karya skripsi ini kepada:

Kedua orang tuaku tercinta, Kukuh Sihana dan Yayuk Sri Rahayu, yang selalu menjadi pelita dalam setiap langkah hidup saya. Terima kasih atas cinta yang tulus, doa yang tak pernah putus, serta pengorbanan yang tiada terhingga demi keberhasilan saya. Setiap tetes keringat dan doa yang kalian panjatkan menjadi kekuatan terbesar saya dalam menghadapi segala tantangan. Semoga karya ini menjadi persembahan kecil atas besarnya cinta dan harapan kalian.

Dosen pembimbing saya, Bapak Dr. Dony Ariyus, S.S., M.Kom. yang dengan sabar membimbing, memberi arahan, dan berbagi ilmu selama proses penyusunan skripsi ini. Terima kasih atas waktu, perhatian, dan dedikasi yang telah diberikan.

Seluruh dosen dan staf pengajar di Program Studi Teknik Komputer yang telah menjadi bagian penting dalam perjalanan akademik saya. Terima kasih atas segala ilmu, motivasi, dan pengalaman berharga yang telah saya dapatkan selama masa perkuliahan.

Teman-teman seperjuangan dan sahabat-sahabat terbaik, yang telah menemani perjalanan panjang ini dengan dukungan, canda tawa, dan kebersamaan yang tak terlupakan. Kalian adalah bagian dari cerita indah dalam hidup saya.

Terakhir, karya ini saya persembahkan kepada Universitas Amikom Yogyakarta, yang telah memberikan ruang untuk tumbuh, belajar, dan bermimpi. Semoga skripsi ini dapat menjadi kontribusi kecil yang bermanfaat dalam pengembangan ilmu pengetahuan, khususnya di bidang keamanan siber dan pembelajaran mesin.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul "Deteksi Website Phishing Berbasis Url Menggunakan Pendekatan Machine learning" ini sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Komputer, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Penyusunan skripsi ini tentu tidak lepas dari bantuan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Bapak Dr. Dony Ariyus, S.S., M.Kom. selaku dosen pembimbing yang telah dengan sabar membimbing, memberikan arahan, serta masukan berharga selama proses penyusunan skripsi ini.
2. Bapak/Ibu dosen dan staf pengajar di Universitas Amikom Yogyakarta yang telah memberikan ilmu dan pengalaman selama masa studi.
3. Kedua orang tua tercinta, Kukuh Sihana dan Yayuk Sri Rahayu, atas segala doa, dukungan moral, dan materiil yang tiada henti.
4. Teman-teman seperjuangan dan seluruh pihak yang tidak dapat penulis sebutkan satu per satu yang telah memberikan bantuan dan semangat.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan untuk perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak yang membacanya, khususnya di bidang Ilmu Komputer.

Yogyakarta, 26 Mei 2025
Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machine
RF	Random Forest
NB	Naive Bayes
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
APJII	Asosiasi Penyelenggara Jasa Internet Indonesia
BSSN	Badan Siber dan Sandi Negara
URL	Uniform Resource Locator
SSL	Secure Socket Layer
DNN	Deep Neural Network
MLP	Multilayer Perceptron
DT	Decision Tree
XGB	Extreme Gradient Boosting
KNN	K-Nearest Neighbors
GNB	Gaussian Naive Bayes
PCA	Principal Component Analysis
SMOTE	Synthetic Minority Oversampling Technique

DAFTAR ISTILAH

Phishing	Teknik serangan siber yang menipu pengguna agar memberikan informasi pribadi.
Machine learning	Teknologi kecerdasan buatan yang memungkinkan komputer belajar dari data.
Deep learning	Cabang machine learning yang menggunakan jaringan saraf dalam (Deep Neural Networks).
Dataset	Kumpulan data yang digunakan untuk melatih dan menguji model machine learning.
Preprocessing	Tahapan persiapan data sebelum pelatihan model, termasuk penghapusan duplikasi, normalisasi, dll.
Feature	Atribut atau karakteristik dari data yang digunakan untuk proses klasifikasi.
Feature selection	Proses pemilihan fitur yang paling relevan untuk meningkatkan akurasi model.
Correlation analysis	Metode statistik untuk mengetahui hubungan antar fitur dalam dataset.
Klasifikasi	Proses mengelompokkan data ke dalam kelas atau kategori tertentu.
Accuracy	Ukuran seberapa sering model memprediksi dengan benar secara keseluruhan.
Precision	Ukuran ketepatan model dalam memprediksi kelas positif.
Recall	Ukuran seberapa banyak data positif yang berhasil dikenali model.
F1-score	Rata-rata harmonis dari precision dan recall.
Confusion matrix	Matriks evaluasi yang menunjukkan jumlah prediksi benar dan salah.
Overfitting	Kondisi ketika model terlalu cocok pada data latih sehingga buruk pada data uji.

Ensemble Learning	Teknik penggabungan beberapa model untuk meningkatkan performa klasifikasi.
SVM	Algoritma klasifikasi yang mencari hyperplane optimal untuk memisahkan kelas.
Random Forest	Algoritma berbasis pohon keputusan yang menggabungkan banyak pohon secara acak.
Naive Bayes	Algoritma klasifikasi probabilistik yang mengasumsikan independensi antar fitur.
SMOTE	Teknik untuk menyeimbangkan dataset dengan membuat sampel minoritas sintetis.
Google Colab	Platform komputasi awan untuk eksperimen Python dan machine learning.
Scikit-learn	Library Python untuk pembelajaran mesin dan analisis data.
Pandas	Library Python untuk manipulasi data tabular.
NumPy	Library Python untuk perhitungan numerik dan operasi array.
Matplotlib	Library visualisasi data dalam bentuk grafik.
Seaborn	Library visualisasi statistik berbasis Matplotlib yang lebih estetis.

INTISARI

Seiring dengan meningkatnya penggunaan internet, ancaman serangan siber seperti *phishing* juga semakin marak terjadi. *Phishing* merupakan metode penipuan yang memanfaatkan teknik rekayasa sosial dan teknologi untuk mencuri informasi sensitif pengguna melalui situs web palsu yang menyerupai situs resmi. Penelitian ini bertujuan untuk mendeteksi situs *phishing* menggunakan pendekatan *machine learning* dengan membandingkan performa tiga algoritma klasifikasi, yaitu *Support Vector Machine* (SVM), *Random Forest* (RF), dan *Naive Bayes* (NB). *Dataset* yang digunakan berasal dari sumber publik di Kaggle dan terdiri dari 11.430 URL yang telah diklasifikasikan sebagai *phishing* dan non-*phishing*. Proses penelitian meliputi tahap *preprocessing* data, seleksi fitur menggunakan analisis korelasi, pembagian data menjadi data latih dan data uji, pelatihan model, serta evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Hasil evaluasi menunjukkan bahwa algoritma *Random Forest* memberikan performa terbaik dengan nilai *accuracy*, *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,96. Algoritma SVM mencatatkan nilai *accuracy* sebesar 0,93, *precision* 0,92, *recall* 0,94, dan *F1-score* 0,93, sedangkan *Naive Bayes* memperoleh *accuracy* sebesar 0,85, *precision* 0,91, *recall* 0,78, dan *F1-score* 0,84. Berdasarkan hasil tersebut, algoritma *Random Forest* dinilai paling efektif untuk diterapkan dalam sistem deteksi *phishing* berbasis URL.

Kata kunci: *Phishing,* Pembelajaran Mesin*, Support Vector Machine, Random Forest, Naive Bayes*

ABSTRACT

With the increasing use of the internet, the threat of cyberattacks such as phishing has also become more widespread. Phishing is a form of fraud that utilizes social engineering techniques and technology to steal sensitive user information through fake websites that resemble legitimate ones. This study aims to detect phishing websites using a machine learning approach by comparing the performance of three classification algorithms: Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB). The dataset used in this research was obtained from a public source on Kaggle and consists of 11,430 URLs classified as either phishing or legitimate. The research process includes data preprocessing, feature selection using correlation analysis, data splitting into training and testing sets, model training, and evaluation using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. The evaluation results show that the Random Forest algorithm achieved the best performance, with accuracy, precision, recall, and F1-score all reaching 0.96. The SVM algorithm achieved an accuracy of 0.93, precision of 0.92, recall of 0.94, and F1-score of 0.93, while Naive Bayes obtained an accuracy of 0.85, precision of 0.91, recall of 0.78, and F1-score of 0.84. Based on these results, the Random Forest algorithm is considered the most effective for implementation in a URL-based phishing detection system.

Keyword: *Phishing, Machine learning, Support Vector Machine, Random Forest, Naive Bayes*