IMPLEMENTASI SISTEM DETEKSI *PHISHING URL* MENGGUNAKAN *MACHINE LEARNING*: *RANDOM FOREST* DAN *BOOSTING* UNTUK PENCEGAHAN *CYBERCRIME*

Disusun Untuk Memenuhi Syarat Menyelesaikan Tugas Akhir
Pendidikan Sarjana Terapan Pada Jurusan Teknik Elektro
Program Studi Teknik Telekomunikasi
Politeknik Negeri Sriwijaya

OLEH:

Aqilla Khairunnisya

062140352366

POLITEKNIK NEGERI SRIWIJAYA

PALEMBANG

2025

DAFTAR TABEL

MOTTO DAN PERSEMBAHAN

"Dan Katakanlah: Bekerjalah kamu, maka Allah akan melihat pekerjaanmu, begitu juga Rasul-Nya dan orang-orang mukmin." (QS. At-Taubah: 105)

"Kemudian apabila engkau telah membulatkan tekad, maka bertawakkallah kepada Allah. Sesungguhnya Allah menyukai orang-orang yang bertawakkal." (QS. Ali 'Imran: 159)

Tugas Akhir ini kupersembahkan kepada:

- ***Allah SWT.***
- ***Ayah, ibu, kakak dan seluruh keluarga besar.***
- ***Ibu Lindawati dan Ibu Suzan Zefi selaku dosen pembimbing, Terima kasih atas bimbingan, arahan, dan dukungan penuh dalam perjalanan tugas akhir ini.***
- ***Teman-teman terdekat yang telah membantu dan mendoakan.***
- ***Teman-teman seperjuangan angkatan 2021***

ABSTRAK

IMPLEMENTASI SISTEM DETEKSI *PHISHING* URL MENGGUNAKAN *MACHINE LEARNING*: *RANDOM FOREST* DAN *BOOSTING* UNTUK PENCEGAHAN CYBERCRIME

(2025:xvi + 65 halaman + 21 gambar + 2 tabel + xx lampiran)

AQILLA KHAIRUNNISYA
062140352366
JURUSAN TEKNIK ELEKTRO
PROGRAM STUDI SARJANA TERAPAN TEKNIK TELEKOMUNIKASI POLITEKNIK NEGERI SRIWIJAYA

Serangan *phishing* melalui URL berbahaya telah menjadi ancaman serius dalam bidang keamanan siber, menyebabkan kerugian finansial yang besar dan kebocoran data secara global. Pendekatan konvensional seperti blacklisting dan deteksi berbasis aturan sering kali tertinggal karena metode *phishing* semakin canggih, termasuk munculnya *phishing* URL zero-day. Dalam penelitian ini, dikembangkan dan diuji model *machine learning* berbasis *Random forest* dan *Gradient boosting* untuk mengidentifikasi *phishing* URL secara akurat. *Dataset* yang digunakan diperoleh dari *Kaggle*, terdiri dari 11.430 URL dengan fitur-fitur yang diekstraksi, mencakup karakteristik seperti panjang URL, jumlah *subdomain*, status HTTPS, dan usia *domain*. Kedua model dilatih dan divalidasi menggunakan skema stratified train-test split serta teknik *cross validation*. Evaluasi performa model dilakukan menggunakan beberapa indikator, antara lain akurasi, presisi, *recall*, *F1-score*, dan ROC AUC. Hasil eksperimen menunjukkan bahwa *Gradient boosting* sedikit melampaui performa *Random forest*, dengan akurasi mencapai 98,0%, presisi 98,1%, dan *F1-score* 98,0%. Model dengan performa terbaik kemudian diintegrasikan ke dalam aplikasi *web* berbasis Streamlit untuk menyediakan deteksi *phishing* secara *real-time* bagi pengguna akhir. Penelitian ini berkontribusi dalam pengembangan sistem deteksi *phishing* URL yang adaptif dan efisien, sehingga dapat meningkatkan pertahanan keamanan siber terhadap ancaman *phishing* yang terus berkembang. Implementasi sistem menunjukkan penerapan yang praktis dan kemudahan penggunaan, bahkan bagi pengguna non-ahli.

Kata kunci: Keamanan Siber, *Gradient boosting*, *Machine learning*, *Phishing* URL, *Random forest*, Streamlit.

ABSTRACT

IMPLEMENTATION OF A PHISHING URL DETECTION SYSTEM USING MACHINE LEARNING: RANDOM FOREST AND BOOSTING FOR CYBERCRIME PREVENTION

(2025:xvi + 65 pages + 21 pictures + 2 tables + xx appendixes)

AQILLA KHAIRUNNISYA
062140352366
ELECTRICAL ENGINEERING DEPARTMENT
PROGRAM OF STUDY IN APPLIED GRADUATION OF THE TELECOMMUNICATION ENGINEERING
STATE POLYTECHNIC OF SRIWIJAYA

Phishing attacks through malicious URLs have become a critical cybersecurity threat, resulting in substantial financial losses and data exposures on a global scale. Conventional approaches like blacklisting and rule-based detection often fall behind as phishing methods become more advanced, including zero-day phishing URLs. In this research, machine learning models based on Random forest and Gradient boosting are designed and tested to accurately identify phishing URLs. The dataset, obtained from Kaggle, consists of 11,430 URLs with extracted features representing URL characteristics such as length, subdomain count, HTTPS status, and domain age. The two models underwent training and validation with the help of stratified train-test splits and cross validation techniques. To evaluate the models, several performance indicators—such as accuracy, precision, recall, F1-score, and ROC AUC—were applied. Results from the experiments reveal that Gradient boosting slightly exceeds the performance of Random forest, achieving an accuracy of 98.0%, precision of 98.1%, and an F1-score of 98.0%. The best-performing model was integrated into a web application built with Streamlit, providing real-time phishing detection for end-users. This research contributes to developing adaptive and efficient phishing URL detection systems, enhancing cybersecurity defenses against evolving phishing threats. The implementation demonstrates practical applicability and ease of use for non-expert users.

Keywords: *Cybersecurity, Gradient boosting, Machine learning, Phising URL, Random forest, Streamlit.*

DAFTAR ISI

LEMBAR PENGESAHAN

IMPLEMENTASI SISTEM DETEKSI *PHISHING* URL MENGGUNAKAN *MACHINE LEARNING*: *RANDOM FOREST* DAN *BOOSTING* UNTUK PENCEGAHAN *CYBERCRIME*

Disusun Untuk Memenuhi Syarat Menyelesaikan Pendidikan Sarjana Terapan Pada Jurusan Teknik Elektro Program Studi Teknik Telekomunikasi Politeknik Negeri Sriwijaya

Oleh:
AQILLA KHAIRUNNISYA
062140352366

Palembang, Agustus 2025

Menyetujui,

Dosen Pembimbing I	Dosen Pembimbing II
Lindawati, S.T., M.TI.	Suzan Zeli, S.T., M. Kom.
NIP 197105282006042001	NIP 197709252009012003

Mengetahui,

Ketua Jurusan Teknik Elektro	Koordinator Program Studi Sarjana Terapan Teknik Telekomunikasi
Dr. Ir. Selamat Muslimin, S.T., M.Kom., IPM.	Mohammad Fadhli, S.Pd., M.T.
NIP 197907222008011007	NIP 199004032018031001

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini menyatakan:

Nama	: Aqilla Khairunnisya
Jenis kelamin	: Perempuan
Tempat, Tanggal Lahir	: Palembang, 17 Februari 2004
Alamat	: Jl. Padang Kapas 2, RT.44 RW 03
NPM	: 062140352366
Program Studi	: Sarjana Terapan Teknik Telekomunikasi
Jurusan	: Teknik Elektro
Judul Skripsi/Laporan Akhir	: Pengembangan Implementasi Sistem Deteksi *Phishing URL* Menggunakan *Machine Learning Random forest* Dan *Boosting* Untuk Pencegahan *Cybercrime*

Menyatakan dengan sesungguhnya bahwa:

1. Skripsi/Laporan Akhir ini adalah hasil karya saya sendiri serta bebas dari tindakan plagiasi dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.
2. Dapat menyelesaikan segala urusan terkait pengumpulan revisi Skripsi/Laporan Akhir yang sudah disetujui oleh dewan penguji paling lama 1 bulan setelah ujian Skripsi/Laporan Akhir.
3. Dapat menyelesaikan segala urusan peminjaman/penggantian alat/buku dan lainnya paling lama 1 bulan setelah ujian Skripsi/Laporan Akhir.

Apabila dikemudian hari diketahui ada pernyataan yang terbukti tidak benar dan tidak dapat dipenuhi, maka saya siap bertanggung jawab dan menerima sanksi tidak diikutsertakan dalam prosesi wisuda serta dimasukkan dalam daftar hitam oleh jurusan Teknik Elektro sehingga berdampak tertundanya pengambilan Ijazah & Transkrip (ASLI&COPY). Demikian surat pernyataan ini dibuat dengan sebenar- benarnya dan dalam keadaan sadar tanpa paksaan.

Palembang, Juli 2025

Yang Menyatakan

(Aqilla Khairunnisya)

DAFTAR GAMBAR