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.*