ABSTRAK

IMPLEMENTASI *NATURAL LANGUAGE PROCESSING* DAN ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK MENDETEKSI URL *PHISHING*

(2025:xv + 60 halaman + 36 gambar +17 tabel + lampiran)

Nabila
062140352380
JURUSAN TEKNIK ELEKTRO
PROGRAM STUDI SARJANA TERAPAN TEKNIK TELEKOMUNIKASI
POLITEKNIK NEGERI SRIWIJAYA

Phishing merupakan bahaya yang signifikan dalam keamanan siber, menggunakan URL jahat untuk menyesatkan pengguna agar mengungkapkan informasi penting. Penelitian ini berupaya untuk membuat model deteksi URL *phishing* menggunakan pembelajaran mesin melalui integrasi ekstraksi fitur URL struktural, metodologi *Natural Language Processing* (NLP), dan algoritma klasifikasi *Support Vector Machine* (SVM). Indikator tren *phishing* diperoleh dari fitur-fitur seperti panjang URL, jumlah titik, dan garis miring, sementara konten URL dikuantifikasi sebagai vektor numerik menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Semua karakteristik selanjutnya diintegrasikan sebagai input ke dalam model *support vector machine* dengan kernel linier untuk klasifikasi. Hasil evaluasi dari laporan klasifikasi menunjukkan bahwa integrasi TF-IDF dan SVM kernel linier mencapai kinerja optimal, dengan akurasi 90%, presisi 92%, *recall* 89%, dan F1-*score* 90%. Sebaliknya, *convusion matrix* menunjukkan akurasi 90,29%, presisi 91,66%, ingatan 88,62%, dan F1-*Score* 90,12%. Studi ini terutama berkontribusi dengan mengintegrasikan NLP dan SVM ke dalam model deteksi phishing adaptif terpadu melalui penggabungan aspek struktural dan tekstual URL. Strategi ini memfasilitasi deteksi *phishing* yang lebih baik dibandingkan dengan teknik yang hanya bergantung pada karakteristik manual. Model ini, tidak seperti penelitian lain yang difokuskan pada contoh tertentu atau NLP yang dikecualikan, dirancang untuk mengidentifikasi banyak kategori URL *phishing* secara luas, sehingga meningkatkan relevansinya dalam menangani serangan siber yang terus berkembang.

Keyword: *Phishing; Natural Language Processing; Support Vector Machine; Term Frequency-Inverse Document Frequency.*

ABSTRACT

IMPLEMENTATION OF NATURAL LANGUAGE PROCESSING AND SUPPORT VECTOR MACHINE ALGORITHM FOR PHISHING URL DETECTION

(2025:xv + 60 *pages* + 36 *pictures* + 17 *tabel* + *appendices*)

Nabila
062140352380
ELECTRICAL ENGINEERING DEPARTMENT
PROGRAM OF STUDY IN APPLIED GRADUATION OF THE TELECOMMUNICATION ENGINEERING
STATE POLYTECHNIC OF SRIWIJAYA

Phishing is a significant threat in cybersecurity, using malicious URLs to deceive users into disclosing sensitive information. This study aims to develop a phishing URL detection model using machine learning through the integration of structural URL feature extraction, Natural Language Processing (NLP) methods, and the Support Vector Machine (SVM) classification algorithm. Phishing trend indicators are obtained from features such as URL length, number of dots, and slashes, while the URL content is quantified as numerical vectors using Term Frequency-Inverse Document Frequency (TF-IDF). All characteristics are integrated as input into a linear kernel support vector machine model for classification. Evaluation results from the classification report indicate that the integration of TF-IDF and linear kernel SVM achieves optimal performance, with 90% accuracy, 92% precision, 89% recall, and a 90% F1-score. Conversely, the confusion matrix shows 90.29% accuracy, 91.66% precision, 88.62% recall, and a 90.12% F1-score. This study contributes significantly by integrating NLP and SVM into a unified adaptive phishing detection model through a combination of URL structural and textual aspects. This strategy enables better phishing detection compared to techniques that rely solely on manual features. Unlike previous studies that focus on specific cases or exclude NLP, this model is designed to identify a wide range of phishing URL categories, enhancing its relevance in addressing evolving cyber threats.

Keywords *: Phishing; Natural Language Processing; Support Vector Machine; Term Frequency-Inverse Document Frequency.*