ABSTRAK

ARYA WINATA. Perbandingan *Kernel Linear* Dan *Kernel Sigmoid* Pada Algoritma SVM Untuk Deteksi *Website Phishing*. Dibimbing oleh Tekad Matulatan dan Berta Erwin SLAM.

Serangan *phishing* yang terus berevolusi melalui manipulasi tautan (URL) secara dinamis menuntut adanya sistem deteksi proaktif untuk menutupi kelemahan metode *blacklist* konvensional yang bersifat reaktif. Penelitian ini bertujuan untuk membandingkan performa algoritma *Support Vector Machine* (SVM) menggunakan fungsi pemetaan Kernel Linear dan Kernel Sigmoid dalam mendeteksi *website phishing*. Sistem deteksi yang dibangun berfokus pada ekstraksi 14 fitur leksikal dari struktur URL. Penelitian ini menggunakan 30.000 sampel URL, yang dikumpulkan dari repositori PhishTank dan Tranco List. Pendekatan pemodelan menerapkan pembagian rasio data 80:20, di mana 24.000 sampel dialokasikan sebagai data latih (*training*) dan 6.000 sampel bertindak sebagai data uji (*testing*). Setelah melalui tahap standardisasi skala (*Feature Scaling*) menggunakan *Z-score*, model dievaluasi dengan menerapkan ambang batas (*threshold*) keputusan probabilitas yang ketat sebesar 0,8 untuk meminimalisir alarm palsu (*False Positive*). Hasil pengujian melalui *K-Fold Cross Validation* (K=5) dan matriks evaluasi menunjukkan bahwa Kernel Linear unggul secara absolut dengan akurasi 98,85%, presisi 99,80%, *recall* 97,90%, dan F1-Score 98,84%, serta komputasi yang cepat (8,73 detik). Sebaliknya, penggunaan kurva non-linier pada Kernel Sigmoid justru mendegradasi performa dengan akurasi 78,60% dan angka kecolongan ancaman (*False Negative*) yang tinggi. Dapat disimpulkan bahwa fitur leksikal URL memiliki korelasi linier yang kuat, sehingga Kernel Linear menjadi pilihan pemetaan yang paling stabil, efisien, dan presisi untuk diimplementasikan ke dalam sistem deteksi keamanan *phishing*.

Kata Kunci: Deteksi *Phishing*, *Support Vector Machine*, Kernel Linear, Kernel Sigmoid, Fitur Leksikal URL.

ABSTRACT

ARYA WINATA. Comparison of Linear Kernel and Sigmoid Kernel in SVM Algorithm for Phishing Website Detection. Supervised by Tekad Matulatan and Berta Erwin SLAM.

Phishing attacks, which continuously evolve through dynamic URL manipulation, demand a proactive detection system to overcome the weaknesses of reactive conventional blacklist methods. This research aims to compare the performance of the Support Vector Machine (SVM) algorithm using Linear Kernel and Sigmoid Kernel mapping functions in detecting phishing websites. The developed detection system focuses on extracting 14 lexical features from the URL structure. This study utilizes 30,000 URL samples collected from the PhishTank and Tranco List repositories. The modeling approach applies an 80:20 data split ratio, where 24,000 samples are allocated as training data and 6,000 samples serve as testing data. After going through the Feature Scaling stage using Z-score standardization, the model is evaluated by applying a strict probability decision threshold of 0.8 to minimize false alarms (False Positives). The evaluation results using K-Fold Cross Validation (K=5) and the confusion matrix show that the Linear Kernel is absolutely superior with an accuracy of 98.85%, a precision of 99.80%, a recall of 97.90%, and an F1-Score of 98.84%, along with fast computation (8.73 seconds). Conversely, the use of the non-linear curve in the Sigmoid Kernel actually degrades performance, resulting in a 78.60% accuracy and a high False Negative rate. It can be concluded that URL lexical features have a strong linear correlation, making the Linear Kernel the most stable, efficient, and precise mapping choice to be implemented into a phishing security detection system.

Keywords: Phishing Detection, Support Vector Machine, Linear Kernel, Sigmoid Kernel, URL Lexical Features.