INTISARI

Ancaman malware terus berkembang dengan tingkat kompleksitas yang semakin tinggi, sementara metode deteksi berbasis *signature* memiliki keterbatasan fundamental dalam menghadapi varian malware baru (*zero-day threats*) yang belum terdaftar dalam basis data, sehingga mendorong kebutuhan akan pendekatan deteksi yang lebih adaptif berbasis analisis konten file secara statistik. Penelitian ini bertujuan membangun sistem deteksi malware berbasis *machine learning* menggunakan algoritma Support Vector Machine (SVM) dengan kernel RBF serta membandingkan performanya terhadap algoritma Random Forest, XGBoost, dan Multi-Layer Perceptron (MLP), dengan pendekatan multi-fitur sebanyak 277 dimensi yang mencakup fitur Shannon entropy, distribusi byte, penanda format binary, enkoding ekstensi file, dan byte histogram penuh, yang diintegrasikan ke dalam aplikasi web berbasis Flask dengan antarmuka *drag-and-drop* untuk pemindaian file secara langsung melalui browser. Pengujian terhadap 193 sampel yang terdiri dari 93 file malware dan 100 file benign menunjukkan bahwa SVM mencapai akurasi 99,48%, *precision* 100%, *recall* 98,92%, dan F1-*Score* 99,46%, mengungguli ketiga algoritma pembanding, di mana nilai *precision* sempurna 100% mengindikasikan tidak ada file benign yang salah diklasifikasikan sebagai malware (*zero false positives*)—karakteristik kritis dalam penerapan sistem deteksi malware secara praktis. Penelitian ini menyimpulkan bahwa kombinasi fitur entropi Shannon dan distribusi byte terbukti efektif sebagai representasi statistik file untuk membedakan malware dari file benign, dan SVM dengan kernel RBF merupakan pilihan algoritma yang optimal untuk klasifikasi biner pada domain deteksi malware.

Kata kunci: deteksi malware, Support Vector Machine, machine learning, Shannon entropy, byte histogram, klasifikasi binary

ABSTRACT

Malware threats continue to grow in complexity, while *signature*-based detection methods have fundamental limitations in identifying new malware variants (*zero-day threats*) that have not yet been registered in existing databases, driving the need for a more adaptive detection approach based on statistical analysis of file content. This research aims to develop a *machine learning*-based malware detection system using the Support Vector Machine (SVM) algorithm with an RBF kernel, and to compare its performance against Random Forest, XGBoost, and Multi-Layer Perceptron (MLP) algorithms, using a 277-dimensional multi-feature approach encompassing Shannon entropy features, byte distribution, binary format markers, file extension encoding, and a full byte histogram, integrated into a Flask-based web application with a *drag-and-drop* interface for direct file scanning through a browser. Evaluation on 193 samples (93 malware and 100 benign files) shows that SVM achieves 99.48% accuracy, 100% *precision*, 98.92% *recall*, and 99.46% F1-*Score*, outperforming all three comparison algorithms, where the perfect 100% *precision* indicates zero false positives—a critical characteristic for practical deployment. This research concludes that the combination of Shannon entropy features and byte distribution proves effective as a statistical file representation for distinguishing malware from benign files, and SVM with an RBF kernel is the optimal algorithm choice for binary classification in the malware detection domain.

Key words: malware detection, Support Vector Machine, machine learning, Shannon entropy, byte histogram, binary classification