ABSTRACT

The reliance of WordPress on third-party plugins results in a high vulnerability rate of 92%; therefore, a machine learning approach is required to automatically identify potential plugin vulnerabilities and enhance the security of the WordPress ecosystem. The dataset used in this study was obtained through a data collection process from the WordPress Plugin API database using a Python script. The modeling process was conducted using three binary classification algorithms, namely Logistic Regression, Naïve Bayes, and Random Forest, implemented using the Orange software. The evaluation method employed a confusion matrix along with evaluation metrics including accuracy, precision, F1-score, and recall. This study demonstrates that the prediction of vulnerable WordPress plugins can be successfully implemented using machine learning algorithms with metadata as predictive features. The implementation process consists of several stages, including Data Collection, Data Preparation, Data Splitting, and Modeling. The evaluation results indicate that all three models achieve good performance, with evaluation scores ranging between 0.75 and 0.80. The results also show that the Random Forest model achieves the best average performance compared to the other models, with scores of 0.797 for accuracy, 0.797 for F1-score, 0.799 for precision, and 0.797 for recall. However, further analysis at the class level reveals that although Random Forest outperforms the other models on most evaluation metrics, there are certain conditions where Logistic Regression provides better results, particularly in terms of precision for the non-vulnerable class and recall for the vulnerable class. This study provides a comprehensive evaluation of each algorithm's performance using standard evaluation metrics and offers recommendations for the most optimal model to be implemented in an automated WordPress plugin vulnerability detection system.

Keywords*: Logistic Regression, Naïve Bayes, Random Forest, WordPress Plugin*

ABSTRAK

Kebutuhan WordPress terhadap plugin pihak ketiga memberikan persentase kerentanan yang tinggi, yaitu 92%, maka diperlukan pendekatan *machine learning* guna mengidentifikasi potensi kerentanan plugin secara otomatis dan meningkatkan keamanan ekosistem WordPress. Dataset yang digunakan dalam penelitian ini diperoleh melalui proses pengumpulan data dari database WordPress Plugin API menggunakan *script* Python. Proses pemodelan dilakukan menggunakan tiga algoritma prediksi biner, yaitu Logistic Regression, Naïve Bayes, dan Random Forest, dengan implementasi menggunakan perangkat lunak Orange. Metode evaluasi menggunakan *confusion matrix* dan matriks evaluasi berupa akurasi, presisi, *F1-score*, dan *recall*. Penelitian ini menunjukkan implementasi prediksi plugin rentan WordPress berhasil dilakukan menggunakan algoritma *machine learning* dengan bahan prediksi berupa metadata. Proses implementasi dimulai dengan tahapan *Data Collection*, *Data Preparation*, Pembagian Data, hingga Pemodelan. Hasil matriks evaluasi menunjukkan ketiga model memiliki performa yang baik dengan rentang skor matriks evaluasi di antara 0.75 – 0.8. Hasil tersebut juga menunjukkan bahwa model Random Forest memiliki performa rerata terbaik dibandingkan model lainnya dengan nilai 0.797 untuk akurasi, 0.797 untuk *F1-score*, 0.799 untuk presisi, dan 0.797 untuk *recall*. Namun, analisis lebih lanjut pada prediksi per kelas menunjukkan bahwa meskipun Random Forest unggul pada sebagian besar metrik evaluasi, terdapat kondisi tertentu di mana Logistic Regression memberikan hasil yang lebih baik, khususnya pada metrik presisi untuk kelas tidak rentan dan *recall* untuk kelas rentan. Penelitian ini menyajikan evaluasi komprehensif terhadap performa masing-masing algoritma menggunakan metrik evaluasi, serta memberikan rekomendasi model yang paling optimal untuk diterapkan dalam sistem deteksi kerentanan plugin WordPress secara otomatis.

Kata Kunci: Logistic Regression, Naïve Bayes, Plugin WordPress, Random Forest