

ABSTRAK

Maraknya pesan *spam* iklan judi *online* pada media sosial WhatsApp yang menggunakan teknik manipulasi teks (*obfuscation*) menjadi ancaman serius terhadap keamanan informasi pengguna. Penelitian ini bertujuan untuk melakukan analisis komparatif serta membandingkan performa algoritma *Naïve Bayes*, *Random Forest*, *Decision Tree*, KNN, *Logistic Regression* dalam mengklasifikasikan pesan *spam* iklan judi *online*. Dataset yang digunakan merupakan data sekunder yang diperoleh dari repositori publik Kaggle dan disaring secara spesifik memuat pesan berbahasa Indonesia. Tahapan penelitian dimulai dari *text preprocessing* (*case folding*, *pembersihan tautan/symbol*, *normalisasi kata leetspeak*, dan *stemming*), ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), hingga pembagian data latih dan data uji sebesar 80:20. Kinerja kedua algoritma dievaluasi dan dibandingkan berdasarkan metrik *Confusion Matrix* yang mencakup *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Hasil analisis komparatif menunjukkan bahwa algoritma *Random Forest* memberikan performa klasifikasi yang lebih superior dan stabil dengan tingkat akurasi sebesar 0.9617%, mengungguli algoritma *Naïve Bayes* yang menghasilkan akurasi sebesar 0.9504%. Pendekatan ensemble pada *Random Forest* terbukti lebih adaptif dalam menangani interaksi antar-fitur dan variasi penulisan teks manipulatif pada karakteristik pesan *spam* judi *online* di WhatsApp.

Kata kunci: Klasifikasi Pesan, *Spam Judi Online*, WhatsApp, *Machine learning*, Analisis Komparatif TF-IDF.

ABSTRACT

The proliferation of online gambling advertisement spam messages on WhatsApp social media utilizing text manipulation techniques (obfuscation) poses a serious threat to user information security. This study aims to conduct a comparative analysis and compare the performance of Naïve Bayes, Random Forest, Decision Tree, KNN, Logistic Regression algorithms in classifying online gambling advertisement spam messages. The dataset used is secondary data obtained from the Kaggle public repository, specifically filtered for Indonesian messages. The research stages include text preprocessing (case folding, link/symbol cleansing, leetspeak word normalization, and stemming), feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF), and data splitting of 80:20 for training and testing data. The performance of both algorithms is evaluated and compared based on Confusion Matrix metrics, including Accuracy, Precision, Recall, and F1-Score. The comparative analysis results indicate that the Random Forest algorithm provides superior and more stable classification performance with an accuracy rate of 0.9617%, outperforming the Naïve Bayes algorithm which yields an accuracy of 0.9504%. The ensemble approach in Random Forest proves to be more adaptive in handling feature interactions and manipulative text variations within the characteristics of online gambling spam messages on WhatsApp.

Keywords: Message Classification, Online Gambling Spam, WhatsApp, Machine learning, Comparative analysis, TF-IDF.