

ABSTRAK

Penelitian ini membandingkan kinerja algoritma Random Forest dan Naïve Bayes dalam mengklasifikasikan pesan WhatsApp ke dalam tiga kategori: normal, promosi, dan penipuan. Dengan lebih dari 2,78 miliar pengguna aktif secara global dan sekitar 90% pengguna internet di Indonesia menggunakan WhatsApp, deteksi spam menjadi tantangan tersendiri karena sistem enkripsi end-to-end yang membatasi penyaringan konten secara tradisional. Oleh karena itu, pendekatan berbasis machine learning menjadi solusi yang relevan. Dataset yang digunakan terdiri dari 300 pesan yang terdistribusi secara seimbang pada masing-masing kategori, dan melalui tahapan pra-pemrosesan seperti pembersihan data, case folding, penghapusan stopword, normalisasi, dan stemming, sebelum diubah ke representasi numerik menggunakan TF-IDF. Hasil pengujian menunjukkan bahwa Naïve Bayes secara konsisten mengungguli Random Forest dalam semua metrik evaluasi, termasuk akurasi (88,67% vs. 86,00%), presisi, recall, dan F1-score. Proses validasi silang *10-fold* juga mengonfirmasi stabilitas performa Naïve Bayes. Selain itu, algoritma ini terbukti lebih efisien secara komputasi, hanya memerlukan 0,13 detik untuk pelatihan dibandingkan 3,65 detik pada Random Forest. Temuan ini memberikan kontribusi penting dalam pengembangan sistem deteksi spam yang efektif pada platform pesan terenkripsi, serta meningkatkan keamanan dan kenyamanan pengguna melalui identifikasi otomatis konten berisiko.

Kata Kunci : *WhatsApp, Klasifikasi Pesan, Naïve Bayes, Random Forest, Text Mining.*

ABSTRACT

This study compares the performance of the Random Forest and Naïve Bayes algorithms in classifying WhatsApp messages into three categories: normal, promotional, and fraudulent. With over 2.78 billion active users globally and approximately 90% of internet users in Indonesia utilizing WhatsApp, spam detection poses a unique challenge due to end-to-end encryption, which restricts traditional content filtering. Therefore, a machine learning-based approach becomes a relevant solution. The dataset consists of 300 messages, evenly distributed across each category, and undergoes preprocessing steps such as data cleaning, case folding, stopword removal, normalization, and stemming before being converted into numerical representations using TF-IDF. The test results demonstrate that Naïve Bayes consistently outperforms Random Forest across all evaluation metrics, including accuracy (88.67% vs. 86.00%), precision, recall, and F1-score. A 10-fold cross-validation process further confirms the stability of Naïve Bayes' performance. Additionally, this algorithm proves to be more computationally efficient, requiring only 0.13 seconds for training compared to 3.65 seconds for Random Forest. These findings contribute significantly to the development of effective spam detection systems on encrypted messaging platforms, enhancing user security and convenience through the automated identification of risky content.

Keywords : *WhatsApp, Message Classification, Naïve Bayes, Random Forest, Text Mining.*