

ABSTRACT

PERBANDINGAN KINERJA ALGORITMA CART, C4.5, DAN C5.0 DALAM KLASIFIKASI EMAIL SPAM DAN HAM (NON-SPAM)

Oleh

YEYEN NOVIA SARI

Email: yeyennovia24@gmail.com

Email merupakan salah satu sarana komunikasi utama di era digital. Namun, seiring perkembangannya, muncul tantangan berupa email spam, yaitu pesan yang tidak diinginkan dan dapat membahayakan pengguna. Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma klasifikasi yaitu CART, C4.5, dan C5.0 dalam mengklasifikasikan email sebagai spam atau ham (non-spam). Data yang digunakan bersumber dari dataset publik berisi 2.638 email berbahasa Inggris yang tersedia di Kaggle. Penelitian dilakukan menggunakan pendekatan kuantitatif dengan tools RapidMiner dan melalui proses text preprocessing serta pembobotan fitur menggunakan TF-IDF. Hasil pengujian menunjukkan bahwa algoritma CART memiliki akurasi tertinggi sebesar 92,29%, disusul oleh C5.0 dengan 91,97%, dan C4.5 dengan 91,53%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma CART paling optimal untuk klasifikasi email spam dan ham dalam konteks data yang digunakan. Penelitian ini memberikan kontribusi dalam pemilihan metode machine learning yang tepat untuk meningkatkan keamanan email pengguna.

Kata Kunci: Email spam, Email ham, Klasifikasi, CART, C4.5, C5.0, RapidMiner.

ABSTRACT

A COMPARATIVE STUDY OF THE PERFORMANCE OF CART, C4.5, AND C5.0 ALGORITHMS IN CLASSIFYING SPAM AND HAM (NON-SPAM)

By :

YEYEN NOVIA SARI

E-mail : yeyennovia24@gmail.com

Email has become one of the primary means of communication in the digital era. However, with its growing use, challenges have emerged, particularly in spam emails-unwanted messages that may pose risks to users. This study aimed to compare the performance of three classification algorithms: CART, C4.5, and C5.0 in categorizing emails as spam or ham (non-spam). The data used was sourced from a public dataset containing 2,638 English-language emails available on Kaggle. The research adopted a quantitative approach using the RapidMiner tool and included text preprocessing and feature weighting using TF-IDF. The experimental results showed that the CART algorithm achieved the highest accuracy at 92.29%, followed by C5.0 at 91.97%, and C4.5 at 91.53%. Based on these results, it can be concluded that the CART algorithm is the most optimal for classifying to choosing more appropriate machine-learning methods for enhancing email security.

Key Words: Email Spam, Email Ham, Classification, CART, C4.5, C5.0, RapidMiner

