ABSTRAK

Meningkatnya penggunaan *email* dalam aktivitas sehari-hari meningkatkan resiko serangan *phishing*, yaitu serangan dalam dunia maya yang dilakukan dengan tujuan untuk memperoleh informasi sensitif dengan berpura-pura menjadi entitas yang sah. *Email phishing* sering kali disamarkan dengan teknik yang sulit dibedakan dari *email* sah. Oleh karena itu diperlukan metode yang efektif untuk mendeteksi *email phishing* secara akurat. Penelitian ini bertujuan untuk menganalisis kinerja algoritma *Random Forest* dalam mendeteksi *email phishing* menggunakan teknik ekstraksi fitur teks *Term Frequency-Inverse Document Frequency* (TF-IDF) serta mengevaluasi kinerja model menggunakan matriks akurasi, presisi, *recall*, dan *F-1 Score*.

Data yang digunakan pada penelitian ini bersumber dari situs *Kaggle* dengan mengambil sampel sebanyak 5000 data *email*. Tahapan *preprocessing* meliputi penggabungan kolom '*subject*' dan '*body*', mengubah kolom 'label' dari numerik menjadi kategorikal, proses *cleaning*, proses *case folding,* proses *tokenization*, proses *stopwords removal*, dan proses *whitespace trimming*. Selanjutnya dilakukan ekstraksi fitur teks menggunakan TF-IDF untuk menghasilkan representasi numerik dari teks pada setiap kata dalam *email* kemudian data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Model klasifikasi dibangun menggunakan algoritma *Random Forest* dengan parameter 50 pohon keputusan dan kedalaman maksimal 10. Evaluasi kinerja model dilakukan berdasarkan hasil prediksi terhadap data pengujian dengan menggunakan *confusion matrix* untuk menghitung akurasi, presisi, *recall*, dan *F-1 Score*.

Hasil pengujian menunjukkan bahwa model *Random Forest* menggunakan TF-IDF yang dikembangkan dapat mendeteksi *email phishing* dengan hasil akurasi sebesar 98%, presisi sebesar 98%, *recall* sebesar 99%, dan *F-1 score* sebesar 98%. Hasil matriks akurasi tersebut menunjukkan bahwa kombinasi TF-IDF dan *Random Forest* efektif untuk klasifikasi *email phishing*. Penelitian ini membuktikan bahwa teknik ekstraksi fitur teks yang tepat dikombinasikan dengan algoritma pembelajaran mesin berbasis *ensemble learning* dapat meningkatkan efektivitas sistem deteksi serangan siber berbasis *email* secara signifikan. Hasil penelitian ini diharapkan dapat menjadi acuan bagi pengembang aplikasi deteksi *email phishing* secara lebih efektif dan efisien.

Kata Kunci: *Email Phishing*, *Random Forest*, TF-IDF, *Confusion Matrix*, Evaluasi Model

ABSTRACT

The increasing use of email in daily activities has raised the risk of phishing attacks, which are cyber-attacks aimed at obtaining sensitive information by pretending to be a legitimate entity. Phishing emails are often disguised using techniques that are difficult to distinguish from legitimate ones. Therefore, an effective method is needed to accurately detect phishing emails. This study aims to analyze the performance of the Random Forest algorithm in detecting phishing emails using the Term Frequency-Inverse Document Frequency (TF-IDF) text feature extraction technique and evaluate the model's performance using accuracy, precision, recall, and F-1 Score metrics.

The data used in this research is sourced from Kaggle, with a sample of 5000 email data. Preprocessing steps include merging the 'subject' and 'body' columns, converting the 'label' column from numeric to categorical, cleaning data, case folding, tokenization, *stopwords removal*, and *whitespace trimming. Text feature extraction using TF-IDF is then performed to generate numerical representations of each word in the email, and the data is split into 80% for training and 20% for testing. A classification model is built using the Random Forest algorithm with 50 decision trees and a maximum depth of 10. The model's performance is evaluated based on the prediction results for the test data, using a confusion matrix to calculate accuracy, precision, recall, and F-1 Score.*

The test results show that the Random Forest model with TF-IDF can detect phishing emails with an accuracy of 98%, precision of 98%, recall of 99%, and F-1 score of 98%. These accuracy matrix results demonstrate that the combination of TF-IDF and Random Forest is effective for phishing email classification. This study proves that the right text feature extraction technique combined with ensemble learning-based machine learning algorithms can significantly improve the effectiveness of email-based cyber attack detection systems. The findings of this study are expected to serve as a reference for developing more effective and efficient phishing email detection applications.

Keywords: *Email Phishing, Random Forest, TF-IDF, Confusion Matrix, Evaluation Model*