ABSTRAK

Identifikasi email phishing merupakan tantangan besar dalam bidang keamanan siber, karena pelaku kejahatan siber terus mengubah taktik mereka untuk mengeksploitasi celah dalam sistem komunikasi. Studi ini mengevaluasi efektivitas model Bidirectional Encoder Representations from Transformers (BERT) dalam mengidentifikasi email phishing, dengan fokus khusus pada pengaruh ukuran dan keragaman dataset. Dua skenario eksperimen dilakukan: Pada Skenario 1, efektivitas BERT diuji menggunakan berbagai dataset email phishing yang unik. Sebaliknya, Skenario 2 menggunakan dataset gabungan yang lebih besar, mencakup 203.176 email. Hasil pada Skenario 1 menunjukkan bahwa BERT melampaui performa model machine learning konvensional seperti SVM, RF, ET, XGB, dan ADB pada berbagai dataset. BERT mencapai akurasi sebesar 99,64% pada dataset Ling, 99,43% pada dataset Enron, dan 99,82% pada dataset TREC-07. Analisis AUC-ROC untuk Skenario 1 juga menunjukkan hasil yang sangat baik, dengan BERT mencapai nilai AUC minimal 99,88% di seluruh dataset. Pada Skenario 2, penggunaan dataset yang lebih besar dan lebih beragam memungkinkan BERT mencapai akurasi sebesar 99,35%, presisi 99,45%, recall 99,04%, F-score 99,24%, serta AUC-ROC sebesar 99,97%. Analisis ini menunjukkan bahwa BERT secara konsisten mengungguli model lain dalam membedakan antara email phishing dan email sah, terlepas dari ukuran dataset. Temuan ini memberikan kontribusi terhadap pengembangan sistem deteksi yang lebih efisien dan memiliki nilai penting dalam memperkuat strategi keamanan siber terhadap serangan phishing di dunia nyata.

Index Terms

Deteksi Phishing Email, Serangan Siber, BERT, Tranformers, Pembelajaran Mendalam, Pembelajaran Mesin