# ABSTRAK

Perkembangan teknologi komunikasi digital yang pesat telah meningkatkan risiko serangan phishing melalui email, sehingga diperlukan metode yang mampu mendeteksi email phishing secara akurat dan andal. Sebelumnya, sistem klasifikasi email phishing banyak menggunakan metode berbasis aturan atau keyword, namun performanya terbatas karena sulit menangani variasi teks yang kompleks dan pola phishing yang terus berkembang. Penelitian ini mengusulkan untuk membuat sistem klasifikasi email phishing dengan membandingkan tiga metode vektorisasi teks yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF), *Count Vectorizer*, dan *Word2Vec* CBOW, menggunakan dua algoritma klasifikasi yaitu Support Vector Machine dan Naive Bayes. Penelitian ini dilakukan untuk menemukan kombinasi metode dan algoritma yang efisien dalam mendeteksi email phishing pada empat macam dataset yaitu CEAS_08, SpamAssasin, Enron, dan Phishing_Email. Hasil penelitian menunjukkan bahwa kombinasi TF-IDF dengan SVM memberikan performa terbaik secara konsisten, dengan akurasi tertinggi pada dataset CEAS_08 sebesar 0,996, diikuti oleh dataset SpamAssassin 0,993, Enron 0,989, dan Phishing_Email 0,977. Hasil ini menegaskan bahwa metode TF-IDF mampu merepresentasikan karakteristik teks email secara efisein, dan SVM memberikan klasifikasi yang stabil serta akurat. Oleh karena itu, kombinasi TF-IDF dan SVM dipilih sebagai metode terbaik untuk sistem deteksi email phishing.

**Kata Kunci:** Email Phishing, Klasifikasi Teks, *Support Vector Machine*, *Naive Bayes*, TF-IDF, *Count Vectorizer*, *Word2Vec*