BAB II

TINJAUAN PUSTAKA

Penggunaan machine learning dengan model algoritma Random Forest dalam mendeteksi URL web phishing telah menjadi fokus dari beberapa penelitian terdahulu. Seperti penelitian dengen judul Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing yang dilakukan oleh Nabila Bianca Putri dan Arie Wahyu Wijayanto[2]. Jurnal ini membahas analisis perbandingan algoritma klasifikasi data mining dalam mengidentifikasi website phishing serta memberikan wawasan yang penting dalam upaya melindungi informasi penting dari serangan phishing melalui penerapan algoritma klasifikasi data mining yang efektif. Penelitian ini menggunakan dataset website phishing sebanyak 1.353 data dengan 10 variabel dan membandingkan kinerja algoritma klasifikasi data mining seperti Naïve Bayes, Random Forest, Decision Tree, dan Support Vector Machine. Hasil penelitian menunjukkan bahwa algoritma Random Forest memiliki kinerja terbaik dengan akurasi sebesar 90,77%, yang lebih tinggi dibandingkan dengan algoritma lainnya. Selain akurasi, presisi dan sensitivitas dari algoritma Random Forest juga menunjukkan nilai tertinggi dibandingkan dengan model algoritma lainnya. Dalam konteks laporan skripsi, penelitian ini dapat menjadi referensi penting dalam mengembangkan metode untuk melindungi informasi penting dari serangan phishing. Dalam laporan skripsi, penelitian ini dapat dijadikan sebagai dasar untuk mengembangkan model klasifikasi yang lebih baik dalam mengidentifikasi website phishing. Selain itu, penelitian ini juga dapat menjadi dasar untuk mengembangkan sistem keamanan yang lebih baik dalam lingkungan online. Dalam laporan skripsi, penelitian ini dapat dijadikan sebagai dasar untuk mengembangkan sistem keamanan yang lebih efektif dalam melindungi pengguna internet dari serangan phishing. Dalam kesimpulannya, penelitian ini memberikan kontribusi penting dalam upaya perlindungan terhadap serangan phishing di lingkungan online. Dengan mengidentifikasi algoritma klasifikasi

terbaik, langkah-langkah perlindungan dapat ditingkatkan untuk mengurangi risiko serangan phishing terhadap pengguna internet. Oleh karena itu, penelitian ini dapat menjadi referensi penting dalam mengembangkan metode dan sistem keamanan yang lebih baik dalam lingkungan online.

Menurut penelitian yang dilakukan oleh Muhammad Fandru Al Rifqi, Mauli Dina, Anita, Marline N.K.Nababan, dan Siti Aisyah mengenai *Comparative Analysis of Phishing Website Prediction Classificationn Algorithm Using Logistic Regression, Decision Tree, and Random Forest*[6]. Penelitian ini penting karena meningkatnya aktivitas sehari-hari yang dilakukan secara daring, yang membuat data pengguna semakin rentan dicuri oleh pihak yang tidak bertanggung jawab. Serangan phishing sendiri merupakan upaya untuk mencuri informasi rahasia dari target dengan mengirimkan tautan palsu. Penelitian ini menggunakan metode analisis data dan pemrosesan menggunakan Python dengan bantuan Google Colab. Algoritma Regresi Logistik, Pohon Keputusan, dan Hutan Acak digunakan untuk melatih dan menguji klasifikasi situs web phishing. Dataset yang digunakan terdiri dari 11054 data situs web dari seluruh dunia dengan total 32 atribut. Penelitian ini melakukan analisis terhadap atribut-atribut tersebut untuk melakukan klasifikasi situs web phishing. Hasil penelitian ini memberikan analisis perbandingan hasil akurasi dari ketiga algoritma yang digunakan. Hasil ini membantu dalam memahami kemampuan masing-masing algoritma dalam memprediksi dan mengklasifikasikan situs web phishing. Jurnal ini memberikan wawasan yang mendalam tentang serangan phishing, metode analisis data, dan perlindungan terhadap serangan phishing. Selain itu, jurnal ini juga memberikan pemahaman yang lebih baik tentang perbandingan algoritma yang digunakan dalam prediksi dan klasifikasi situs web phishing. Dengan demikian, jurnal ini memberikan kontribusi dalam upaya meningkatkan keamanan internet dan melindungi data pengguna dari serangan phishing. Jurnal ini dapat menjadi referensi yang berharga untuk mendukung penelitian tentang keamanan internet, analisis data, dan perlindungan terhadap serangan phishing.

Pada penelitian dengan judul Identifikasi *Website Phishing* dengan Perbandingan Algoritma Klasifikasi, yang dilakukan oleh Agung Susilo Yuda Irawan, Nono Heryana, Hopi Siti Hopipah, dan Dyas Rahma Putri[7], membahas tentang phishing yang merupakan salah satu jenis kejahatan siber yang mengancam dan menjebak seseorang dengan cara memancing korban untuk memberikan informasi kepada penjebak. Oleh karena itu, penting untuk dapat mendeteksi website phishing agar dapat menghindari kerugian yang mungkin ditimbulkan. Penelitian ini dilakukan dengan menggunakan dataset publik yang terdiri dari tiga kategori website: legitimate, suspicious, dan phishing. Penelitian ini membandingkan performa empat algoritma klasifikasi, yaitu Support Vector Machine (SVM), Decision Tree, Random Forest, dan Multilayer Perceptron. Hasil penelitian menunjukkan bahwa algoritma Multilayer Perceptron memiliki performa terbaik dalam mendeteksi website phishing, dengan tingkat akurasi mencapai 93.15% dan nilai AUC 0.976. Peneliti juga menyoroti potensi kerugian yang dapat ditimbulkan oleh website phishing bagi korban, seperti kerugian dalam hal privacy, eksploitasi data, dan kerugian finansial. Oleh karena itu, identifikasi website phishing menjadi sangat penting untuk menghindari kerugian tersebut. Dalam penelitian ini, evaluasi performa algoritma klasifikasi dilakukan dengan menggunakan parameter tingkat akurasi, precision, recall, F1-score, dan nilai AUC. Dokumen ini memberikan wawasan tentang pentingnya identifikasi website phishing dan perbandingan performa algoritma klasifikasi dalam mendeteksinya.