ABSTRAK

NANDA ALIF BAKRI. Analisis Perbandingan Algoritma *LightGBM* dan *Random Forest* dalam Mendeteksi *Phishing* Berbasis *URL*. Dibimbing oleh TEKAD MATULATAN dan FERI IRAWAN.

Phishing merupakan salah satu bentuk kejahatan siber yang bertujuan untuk mencuri informasi sensitif pengguna melalui tautan atau situs *web* tiruan yang menyerupai layanan resmi. Serangan *phishing* berbasis *URL* menjadi salah satu bentuk serangan yang sering digunakan karena pelaku dapat membuat struktur *URL* yang terlihat mirip dengan *website* asli sehingga sulit dikenali secara langsung oleh pengguna. Untuk mengatasi permasalahan tersebut, diperlukan sistem deteksi yang mampu mengenali pola *URL phishing* secara otomatis. Penelitian ini membandingkan dua algoritma *machine learning*, yaitu *LightGBM* dan *Random Forest* dalam mendeteksi *URL phishing* dan *legitimate*. Dataset yang digunakan berasal dari *Kaggle* tahun 2021 dengan jumlah 11.340 *URL*. Hasil pengujian menunjukkan bahwa model *LightGBM* memiliki performa yang sedikit lebih tinggi dibandingkan *Random Forest*, di mana *LightGBM* mencapai *accuracy* sebesar 86,09%, *precision* 88,06%, *recall* 83,49%, dan *F1-score* 85,71%. Sementara itu, algoritma *Random Forest* memperoleh *accuracy* sebesar 85,56%, *precision* 86,61%, *recall* 84,13%, dan *F1-score* 85,35%. Temuan ini menunjukkan bahwa kedua model mampu melakukan deteksi *phishing* berbasis *URL* dengan cukup baik, namun algoritma *LightGBM* lebih unggul secara keseluruhan pada *accuracy*, *precision*, dan *F1-score*, sedangkan *Random Forest* memiliki nilai *recall* yang sedikit lebih tinggi.

Kata kunci: *phishing*, *URL*, *machine learning*, *LightGBM*, *Random Forest*.

ABSTRACT

NANDA ALIF BAKRI. Comparative Analysis of LightGBM and Random Forest Algorithms in URL-Based Phishing Detection. Supervised by TEKAD MATULATAN and FERI IRAWAN.

Phishing is a form of cybercrime that aims to steal users' sensitive information through fake links or websites designed to resemble legitimate services. URL-based phishing attacks are commonly used because attackers can create URL structures that appear similar to legitimate websites, making them difficult for users to identify directly. To address this problem, a detection system capable of automatically recognizing phishing URL patterns is required. This study compares two machine learning algorithms, namely LightGBM and Random Forest, in detecting phishing and legitimate URLs. The dataset used in this study was obtained from Kaggle in 2021 and consists of 11,340 URLs. The evaluation results show that the LightGBM model achieved slightly better performance than Random Forest, with an accuracy of 86.09%, precision of 88.06%, rsecall of 83.49%, and F1-score of 85.71%. Meanwhile, the Random Forest algorithm achieved an accuracy of 85.56%, precision of 86.61%, recall of 84.13%, and F1-score of 85.35%. These findings indicate that both models are capable of detecting URL-based phishing with reasonably good performance. However, LightGBM performs better overall in terms of accuracy, precision, and F1-score, while Random Forest achieves a slightly higher recall value.

Keywords: phishing, URL, machine learning, LightGBM, Random Forest.