ABSTRAK

Tautan keamanan web makin bertambah setiap harinya, dan deteksi serta analisis halaman web yang berbahaya makin penting dan sulit dilakukan, *Phishing* adalah salah satu bentuk kejahatan dengan cara mendapatkan informasi sensitif dengan meniru pengirim yang terpercaya di dalam saluran komunikasi jika dalam dunia internet berarti meniru *domain name* dari sebuah *website* sehingga pengguna internet tertipu dan akhirnya memberikan data pribadi mereka.

Dikarenakan perkembangan teknologi dan membludaknya *URL* (tautan) di internet maka kita tidak dapat melakukan pengecekan satu persatu terhadap tautan tautan yang ada, maka dibangunlah sistem untuk membantu dalam hal tersebut.

Prediksi *URL* berbahaya dibangun menggunakan *Deep Learning* dengan metode campuran *Long Short term Memory* (LSTM) dan *Convolutional Neural Network* (CNN). Pelatihan dan pengujian model tersebut dilakukan pada platform *Kaggle Notebook.* Setelah selesai, model akan dievaluasi akurasinya dengan menggunakan K-Fold Cross Validation.

Dalam penelitian ini arsitektur model yang mendapatkan akurasi terbaik adalah arsitektur 4 layer LSTM dengan akurasi 92,990%, pengabungan layer CNN sebelum LSTM dapat memperpendek waktu pelatihan model dari 119.45 menit ke 30.86 menit walaupun menghasilkan akurasi lebih rendah di 92,302%.

Kata kuci : tautan berbahaya, *neural network, phishing, fraud detection, kaggle notebook.*

ABSTRACT

Web security links are increasing every day, and detecting and analyzing malicious web pages is becoming increasingly important and difficult to do. Phishing is a form of crime in which sensitive information is obtained by impersonating a trusted sender in a communication channel. In the internet world, this means imitating the domain name of a website so that internet users are deceived and end up giving their personal data.

Due to technological developments and the explosion of URLs (links) on the internet, it is not possible to check each link one by one. Therefore, a system has been built to assist in this matter.

The prediction of malicious URLs is built using Deep Learning with a mixed method of Long Short-term Memory (LSTM) and Convolutional Neural Network (CNN). Training and testing of the model is done on the Kaggle Notebook platform. After completion, the model will be evaluated for accuracy using K-Fold Cross Validation.

In this study, the model architecture that obtains the best accuracy is the 4-layer LSTM architecture with an accuracy of 92.990%. Combining the CNN layer before LSTM can shorten the model training time from 119.45 minutes to 30.86 minutes, although it results in lower accuracy at 92.302%.

Keywords: malicious links, neural network, phishing, fraud detection, kaggle notebook.