IMPLEMENTASI EKSTENSI GOOGLE CHROME DALAM MENDETEKSI SITUS WEB *PHISHING* MENGGUNAKAN ALGORITMA *RANDOM FOREST*

MUHAMAD ABDUL GHANNI AL GHIFARI

ABSTRAK

Di zaman teknologi ini penggunaan internet mengubah kehidupan sehari-hari masyarakat menjadi lebih mudah. Dengan segala manfaat internet terdapat oknum tertentu yang melakukan berbagai macam kejahatan yaitu salah satunya membuat situs web *phishing*. Banyak korban tanpa sadar memasukan data rahasianya ke situs web *phishing* karena URL domain, konten, tampilan dan lain-lain dibuat semirip mungkin dengan situs web aslinya. Karena hal tersebut penelitian ini mengusulkan pendeteksian situs web *phishing*. Untuk mendeteksi situs web *phishing* pendekatan yang paling umum digunakan yaitu metode *blacklist* dan *whitelist,* namun metode ini mempunyai beberapa kekurangan yakni tidak semua URL atau URL yang baru dibuat langsung ada pada *database* tersebut. Oleh sebab itu penelitian ini bertujuan menggunakan pendekatan *machine learning* yaitu metode *Random Forest,* dengan mengimplementasikannya ke dalam ekstensi peramban seperti Google Chrome. Ekstensi peramban ini mengekstrak fitur yang ada pada URL dan halaman situs web saja, tanpa bergantung kepada *web service*, supaya pendeteksian dapat lebih cepat. Hasil evaluasi model klasifikasi mempunyai hasil akurasi 90,2%, recall 88,8% dan presisi 88,8%. Setelah model diimplementasikan ke ekstensi peramban dilakukan evaluasi kinerja menggunakan data baru dengan akurasi 88%, *recall* 84% dan presisi 91,3%, yang mana mengalami penurunan kinerja tetapi masih cukup baik dibandingkan dengan deteksi situs web *phishing default* dari *Google Safe Browsing* (GSB) pada Google Chrome yang mempunyai rata-rata akurasi ~45%.

Kata kunci : Deteksi *Phishing*, *Machine Learning*, Klasifikasi, *Random Forest*, Ekstensi Peramban.

IMPLEMENTATION OF GOOGLE CHROME *EXTENSION IN DETECTING PHISHING WEBSITES USING RANDOM FOREST ALGORITHM*

MUHAMAD ABDUL GHANNI AL GHIFARI

ABSTRACT

In this technological age, the use of the internet has made people's daily lives easier. With all the benefits of the internet, there are certain people who commit various kinds of crimes, one of which is creating phishing websites. Many victims unknowingly enter their confidential data into phishing websites because the domain URL, content, appearance and others are made as similar as possible to the original website. Because of this, this research proposes phishing website detection. To detect phishing websites, the most commonly used approach is the blacklist and whitelist method, but this method has some drawbacks, namely not all URLs or newly created URLs are directly in the database. Therefore, this research aims to use a machine learning approach, namely the Random Forest method, by implementing it into a browser extension such as Google Chrome. This browser extension extracts features from URLs and web pages only, without relying on web services, so that detection can be faster. The classification model evaluation results have an accuracy of 90.2%, recall 88.8% and precision 88.8%. After the model was implemented into the browser extension, a performance evaluation was carried out using new data with an accuracy of 88%, recall 84% and precision of 91.3%, which decreased performance but was still quite good compared to the default phishing website detection from Google Safe Browsing (GSB) on Google Chrome, which has an average accuracy of ~45%.

Keywords: *Phishing Detection, Machine Learning, Classification, Random Forest, Browser Extension.*