INTISARI

Pendeteksian Phishing Website berdasarkan Alamat Website dengan Metoda Recurrent Neural Network

Oleh

Faridin

22/510916/PPA/06487

Phishing merupakan salah satu teknik *social engineering* yang efektif dalam aksi *cybercrime*, karena memiliki tingkat keberhasilan yang tinggi. Teknik ini dapat memancing rasa keingintahuan ataupun memanfaatkan ketidaktahuan target, untuk memperoleh informasi personal, akses maupun menyusupkan *malware*. *Phishing website* merupakan salah satu jenis *phishing* yang memanfaatkan *website* dalam melakukan aksinya, target akan diarahkan secara langsung(*direct*) atau tidak langsung(*indirect*) ke suatu *phishing website*, kemudian memancing target untuk memberikan informasi maupun akses yang minta secara sadar maupun tidak sadar.

Saat ini telah berkembang berbagai upaya memitigasi serangan *phishing website* mulai dari pedekatan yang berbasis daftar (*blacklist/white listing based*), analisa *heuristic* (analisa *content*), maupun melalui pendekatan *machine learning*. Pendekatan *machine learning* dapat memberikan efektivitas pendeteksian yang lebih baik dari pendekatan lainnya, dengan akurasinya mencapai hingga 98%, namun demikian saat ini juga berkembang pendekatan *deep learning*, yang mampu memberikan tingkat akurasi lebih baik dari *machine learning*.

Pada penelitian ini melakukan obersevasi efektivitas dari pendekatan *deep learning* yang menggunakan algoritma RNN: *Gate Recurrent Unit* (GRU) dan *Long Short Time Memory* (LSTM), terhadap pendekatan *machine learning* yang menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF). dan observasi efesiensi model *Recurrent Neural Network* (RNN) dengan merujuk pada penelitian Feng & Yue (2020) sebagai basis.

Hasil penelitian untuk efektifitas pendeteksian *phishing website* dengan pendekatan *deep learning* RNN diperoleh hasil yang sangat efektif dengan tingkat akurasi mencapai hingga 99%. Selain itu dengan metoda *deep learning* tidak memerlukan analisa leksikal, dan proses fiturisasi yang rumit seperti pada *machine learning*. Sedangkan untuk efisiensi model *deep learning* RNN diperoleh konfigurasi model yang lebih efisien dari model *deep learning* RNN rujukan sebelumnya, yaitu waktu *training* model yang rata-rata lebih cepat 75%, dan waktu inferensi rata-rata yang lebih cepat 2%.

Kata Kunci: URL Phishing, Phishing website detection, GRU, LSTM, RNN

Abstract

Phishing is a highly effective social engineering technique used in cybercrime to exploit curiosity or lack of awareness, leading to personal information theft or malware introduction. Phishing websites specifically direct targets to malicious sites to extract information or access, often without their knowledge. Various mitigation strategies exist, including blacklist/whitelist approaches, heuristic analysis, and machine learning, with machine learning offering detection accuracies up to 98%. However, deep learning approaches, such as Recurrent Neural Networks (RNN), promise even higher accuracy.

This research evaluates deep learning RNN algorithms, specifically Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM), against machine learning approaches such as Support Vector Machine (SVM) and Random Forest (RF). It also assesses the efficiency of RNN models by referencing previous research results.

The research results indicate that deep learning RNN approaches are highly effective for phishing website detection, achieving up to 99% accuracy. Additionally, deep learning simplifies the complex extraction and feature selection processes of machine learning with a straightforward tokenization process. The study shows that the RNN model is more efficient than previous models, with a 75% reduction in average training time and a 2% improvement in inference time.

Keywords— *URL Phishing, Phishing website detection, GRU, LSTM, RNN*