22/S/TEKKOM-KCBR/PK.03.08/23/JULI/2024

IMPLEMENTASI ALGORITMA *LONG SHORT-TERM MEMORY* PADA SISTEM DETEKSI SITUS *PHISHING* BERBASIS APLIKASI *WEB*

SKRIPSI

diajukan untuk memenuhi sebagian syarat
untuk memperoleh gelar Sarjana Teknik pada Program Studi Teknik Komputer

oleh
Anisa Nur Syafia
NIM 2008567

PROGRAM STUDI TEKNIK KOMPUTER
KAMPUS UPI DI CIBIRU
UNIVERSITAS PENDIDIKAN INDONESIA

2024

HALAMAN HAK CIPTA

IMPLEMENTASI ALGORITMA *LONG SHORT-TERM MEMORY* PADA SISTEM DETEKSI SITUS *PHISHING* BERBASIS APLIKASI *WEB*

oleh
Anisa Nur Syafia
NIM 2008567

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana Teknik pada Program Studi Teknik Komputer

HALAMAN PENGESAHAN SKRIPSI

ANISA NUR SYAFIA

IMPLEMENTASI ALGORITMA *LONG SHORT-TERM MEMORY* PADA SISTEM DETEKSI SITUS *PHISHING* BERBASIS APLIKASI *WEB*

disetujui dan disahkan oleh pembimbing:

Pembimbing I

Deden Pradeka, S.T., M.Kom.
NIP. 920200419890816101

Pembimbing II

Muhammad Taufik Dwi Putra, S.Tr.Kom., M.T.I.
NIP. 920200819940117101

Mengetahui,
Ketua Program Studi Teknik Komputer

Deden Pradeka, S.T., M.Kom.
NIP. 920200419890816101

HALAMAN PERNYATAAN
KEASLIAN SKRIPSI DAN BEBAS PLAGIARISME

Dengan ini saya menyatakan bahwa skripsi dengan judul "Implementasi Algoritma *Long Short-Term Memory* pada Sistem Deteksi Situs *Phishing* berbasis Aplikasi *Web*" ini beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Bandung, Juli 2024

Penulis,

Anisa Nur Syafia

NIM. 2008567

HALAMAN UCAPAN TERIMA KASIH

Puji serta syukur penulis panjatkan kepada Allah SWT yang telah melimpahkan rahmat, taufiq dan hidayat-Nya sehingga penulis dapat menyelesaikan penelitian dan menghasilkan skripsi dengan judul "Implementasi Algoritma *Long Short-Term Memory* pada Sistem Deteksi Situs *Phishing* berbasis Aplikasi *Web*". Penyusunan skripsi ini merupakan sebagian dari syarat kelulusan untuk memperoleh gelar Sarjana Teknik pada program studi Teknik Komputer, Universitas Pendidikan Indonesia. Dalam proses penyusunan hingga penyelesaian skripsi ini tentunya tidak terlepas dari bantuan, bimbingan dan dukungan dari berbagai pihak. Maka dari itu, penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT dengan segala rahmat serta karunia-Nya yang selalu memberikan kesehatan, kemudahan, kelancaran, dan kekuatan kepada penulis sehingga dapat menyelesaikan penelitian ini.
2. Kedua orang tua penulis, Bapak Asep Kurnia dan Ibu Masnia Fatimah yang selalu memberikan doa dan dukungan tiada henti.
3. Keluarga penulis, Andita Nurfaridah, Andini Syafa Fauziyya, Ananda Zahira, Andhika Sajidan, Aksara Ziyech Ramdani dan Aruna yang selalu menjadi motivasi bagi penulis.
4. Bapak Deden Pradeka, S.T., M.Kom., selaku Ketua Program Studi Teknik Komputer Universitas Pendidikan Indonesia Kampus Cibiru dan Dosen Pembimbing I yang bersedia meluangkan waktunya untuk memberikan ilmu dan bimbingan kepada penulis.
5. Bapak Muhammad Taufik Dwi Putra, S.Tr.Kom., M.T.I., selaku Dosen Pembimbing II yang bersedia meluangkan waktunya untuk memberikan ilmu dan bimbingan kepada penulis.
6. Bapak Wirmanto Suteddy, S.T., M.T., selaku Dosen Pembimbing Akademik yang telah membimbing dan memberikan saran selama masa perkuliahan.
7. Seluruh Bapak dan Ibu dosen serta staf Program Studi Teknik Komputer Universitas Pendidikan Indonesia Kampus Cibiru yang telah membimbing dan memberikan saran selama masa perkuliahan.

8. Aisyah Amatullah, Sofia Marsha Ramadani, Dinda Ilhan Pujihandayani, Anisya Putri Saskia yang selalu membersamai penulis dan memberikan motivasi.
9. Nazwa Putri Nadhipa, Tiara Afriani, Rahmawati, Aulia Putri Cendikia, Syiva Awaliyah Maqdis dan Naziva Septian yang selalu membersamai penulis dan memberikan motivasi.
10. Kak Rial yang selalu membersamai penulis dan memberikan motivasi.
11. Teman-teman Teknik Komputer angkatan 2020 yang membersamai penulis dalam menjalani masa perkuliahan.
12. Teman-teman dan pihak lain yang tidak dapat penulis sebutkan satu per satu yang telah mendoakan, memberikan bantuan dan dukungan dalam proses penyelesaian skripsi ini.

Penulis berharap hasil dari penelitian dapat memberikan manfaat di bidang penerapan ilmu pengetahuan dalam hal keamanan data, yaitu pencegahan serangan situs *phishing*. Penulis menyadari masih banyak kekurangan dalam penelitian ini, maka dari itu penulis mengharapkan kritik dan saran yang membangun untuk perbaikan yang lebih baik lagi. Akhir kata penulis ucapkan terima kasih sekali lagi kepada pihak yang telah membantu dan semoga penelitian ini memberikan manfaat bagi banyak orang.

Bandung, Juli 2024

Penulis,

Anisa Nur Syafia

NIM. 2008567

IMPLEMENTASI ALGORITMA *LONG SHORT-TERM MEMORY* PADA SISTEM DETEKSI SITUS *PHISHING* BERBASIS APLIKASI *WEB*

Anisa Nur Syafia

2008567

ABSTRAK

Serangan situs *phishing* merupakan tindakan kejahatan di dunia maya yang melibatkan usaha untuk mencuri identitas pengguna melalui situs *web* palsu dengan menyamar sebagai lembaga sah yang dapat dipercaya. Deteksi situs *phishing* berbasis algoritma *Deep Learning* diperlukan untuk menghindari serangan *phishing* yang dapat membahayakan keamanan data. Implementasi algoritma *Long Short-Term Memory* (LSTM) pada sistem deteksi situs *phishing* merupakan penelitian yang berfokus pada pengembangan model dan aplikasi *web* sebagai sistem deteksi situs *phishing*. Untuk membangun sebuah model algoritma LSTM, keakuratan deteksi sangat bergantung terhadap jumlah data yang digunakan dalam proses pelatihan dan pengujian juga pemahaman antar fitur datanya. Jumlah data yang digunakan dalam penelitian ini sebanyak 784.762 URL dengan pembagian data diantaranya 392.381 URL *phishing* dan 392.381 URL sah. Dengan memanfaatkan penggunaan klasifikasi menggunakan algoritma LSTM, model dapat memahami tiap karakter fiturnya tanpa ekstraksi fitur manual dan menghasilkan nilai akurasi yang baik yaitu sebesar 98,87%. Model yang dihasilkan kemudian diimplementasikan pada aplikasi *web* yang dibangun menggunakan *framework flask*. Berdasarkan hasil pengujian *black box*, aplikasi *web* yang dibangun dapat berfungsi dengan optimal karena dapat menjalankan semua fitur aplikasi dengan baik dan sesuai. Aplikasi *web* yang diintegrasikan dengan model LSTM juga berhasil membedakan URL sah dan *phishing* dengan cepat dan akurat sehingga dapat meningkatkan keamanan data dan melindungi pengguna internet dari serangan situs *phishing* yang terus berkembang.

Kata KuncI: LSTM, *phishing*, aplikasi *web*, *flask*, *black box*

IMPLEMENTATION OF THE LONG SHORT-TERM MEMORY ALGORITHM IN A WEB APPLICATION-BASED PHISHING SITE DETECTION SYSTEM

Anisa Nur Syafia
2008567

ABSTRACT

A phishing website attack is a cybercrime that involves attempting to steal a user's identity through a fake website by posing as a legitimate, trustworthy institution. Deep Learning algorithm-based phishing site detection is needed to avoid phishing attacks that can endanger data security. The implementation of the Long Short-Term Memory (LSTM) algorithm in a phishing site detection system is research that focuses on developing models and web applications as a phishing site detection system. To build an LSTM algorithm model, detection accuracy is very dependent on the amount of data used in the training and testing process as well as understanding between data features. The amount of data used in this research was 784,762 URLs with data distribution including 392,381 phishing URLs and 392,381 legitimate URLs. By utilizing classification using the LSTM algorithm, the model can understand each character feature without manual feature extraction and produces a good accuracy value of 98.87%. The resulting model is then implemented in a web application built using the flask framework. Based on the results of black box testing, the web application that is built can function optimally because it can run all the application features properly and appropriately. Web applications integrated with the LSTM model also succeed in distinguishing between legitimate and phishing URLs quickly and accurately, thereby increasing data security and protecting internet users from attacks by phishing sites that continue to grow.

Keywords: *LSTM, phishing, web applications, flask, black box*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR PUSTAKA

Aldakheel, E. A., Zakariah, M., Gashgari, G. A., Almarshad, F. A., & Alzahrani, A. I. A. (2023). A Deep Learning-Based Innovative Technique for Phishing Detection in Modern Security with Uniform Resource Locators. *Sensors*, *23*(9), 4403. https://doi.org/10.3390/s23094403

Almousa, M., Zhang, T., Sarrafzadeh, A., & Anwar, M. (2022). Phishing website detection: How effective are deep learning-based models and hyperparameter optimization? *Security And Privacy*, *5*(6). https://doi.org/10.1002/spy2.256

Anti-Phishing Working Group (APWG). (2023). *Phishing Activity Trends Report*. http://www.apwg.org,

Atawneh, S., & Aljehani, H. (2023). Phishing Email Detection Model Using Deep Learning. *Electronics*, *12*(20), 4261. https://doi.org/10.3390/electronics12204261

Badan Siber dan Sandi Negara (BSSN). (2023). *Lanskap Keamanan Siber Indonesia 2023*. https://www.bssn.go.id/wp-content/uploads/2024/03/Lanskap-Keamanan-Siber-Indonesia-2023.pdf

Bahnsen, A. C., Bohorquez, E. C., Villegas, S., Vargas, J., & Gonzalez, F. A. (2017). Classifying phishing URLs using recurrent neural networks. *2017 APWG Symposium on Electronic Crime Research (eCrime)*, 1–8. https://doi.org/10.1109/ECRIME.2017.7945048

Do, N. Q., Selamat, A., Krejcar, O., Yokoi, T., & Fujita, H. (2021). Phishing Webpage Classification via Deep Learning-Based Algorithms: An Empirical Study. *Applied Sciences*, *11*(19), 9210. https://doi.org/10.3390/app11199210

Ghifari, M. A. G. Al, Hananto, B., & Wahyono, B. T. (2022). Implementasi Ekstensi Google Chrome Dalam Mendeteksi Situs Web Phishing Menggunakan Algoritma Random Forest. *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, *3*(2), 640–649.

Hannousse, A., & Yahiouche, S. (2021). Towards benchmark datasets for machine learning based website phishing detection: An experimental study. *Engineering Applications of Artificial Intelligence*, *104*(104347). https://doi.org/10.1016/j.engappai.2021.104347

Liu, D. J., Geng, G. G., Jin, X. B., & Wang, W. (2021). An efficient multistage phishing website detection model based on the CASE feature framework: Aiming at the real web environment. *Computers and Security*, *110*. https://doi.org/10.1016/j.cose.2021.102421

Muhyidin, H. A. F., & Venica, L. (2023). Pengembangan Chatbot untuk Meningkatkan Pengetahuan dan Kesadaran Keamanan Siber Menggunakan

Long Short-Term Memory. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, *5*(2), 152–161. https://doi.org/10.36499/jinrpl.v5i2.8818

Nawali, I., & Suteja, B. R. (2023). Pembuatan Sistem Aplikasi Berbasis Website Konsultasi Orang Tua dengan Psikolog untuk Kesehatan Mental Anak. *Jurnal Strategi*, *5*(1), 110–129.

Nugraha, A. F., Aziza, R. F. A., & Pristyanto, Y. (2022). Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing. *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan* , *7*(1), 39–44.

Perbawa, D. S., & Nurohim, G. S. (2020). Pengujian Aplikasi Berbasis Website Dengan Black Box Testing Metode Boundary Value Analysis Dan Responsive Testing. *Journal Speed-Sentra Penelitian Engineering dan Edukasi*, *12*(4).

Prasad, A., & Chandra, S. (2024). *PhiUSIIL Phishing URL (Website)*. UCI Machine Learning Repository. https://doi.org/10.1016/j.cose.2023.103545

Putra, M. T. D., Ardimansyah, M. I., & Aprianti, D. (2022). Deteksi Konten Pornografi Menggunakan Convolutional Neural Network Untuk Melindungi Anak Dari Bahaya Pornografi. *Jurnal Media Informatika Budidarma*, 6(4), 2401–2409. https://doi.org/10.30865/mib.v6i4.4793

Rahman, S. S. M. M., Islam, T., & Jabiullah, Md. I. (2020). PhishStack: Evaluation of Stacked Generalization in Phishing URLs Detection. *Procedia Computer Science*, *167*, 2410–2418. https://doi.org/10.1016/j.procs.2020.03.294

Rajeswary, C., & Thirumaran, M. (2023). The LSTM-based automated phishing detection driven model for detecting multiple attacks on Tor hidden services. *Journal of Intelligent & Fuzzy Systems*, *44*(6), 8889–8903. https://doi.org/10.3233/JIFS-224142

Rifnaldy, R., & Tony. (2023). Perancangan Aplikasi Media Informasi Dan Pemesanan Berbasis Web Untuk Coffee Shop Tempat Bercerita. *Jurnal Ilmu Komputer dan Sistem Informasi*, *11*(1).

Somesha, M., Pais, A. R., Rao, R. S., & Rathour, V. S. (2020). Efficient deep learning techniques for the detection of phishing websites. *Sādhanā*, *45*(1), 165. https://doi.org/10.1007/s12046-020-01392-4

Sudhan, H. (2024). *Phishing and Legitimate URLS*. Kaggle. https://www.kaggle.com/datasets/harisudhan411/phishing-and-legitimate-urls/data

Suraya, & Sholeh, M. (2021). Designing and Implementing a Database for Thesis Data Management by Using the Python Flask Framework. International Journal of Engineering, Science and Information Technology, 2(1), 9–14. https://doi.org/10.52088/ijesty.v2i1.197

Wahyudi, D., Niswar, M., & Alimuddin, A. A. P. (2022). Website Phising Detection Application Using Support Vector Machine (SVM). *Journal Of Information Technology And Its Utilization*, *5*(2).

Wiharja, S. A. J., Pradeka, D., & Suteddy, W. (2024). Comparative Study of the Effect of Datasets and Machine Learning Algorithms for PDF Malware Detection. Digital Zone: Jurnal Teknologi Informasi dan Komunikasi, 15(1), 80–93.

Windarni, V. A., Nugraha, A. F., Ramadhani, S. T. A., Istiqomah, D. A., Puri, F. M., & Setiawan, A. (2023). Deteksi Website Phishing Menggunakan Teknik Filter Pada Model Machine Learning. *Information System Journal (INFOS)*, *6*(1), 39–43.

Yang, P., Zhao, G., & Zeng, P. (2019). Phishing website detection based on multidimensional features driven by deep learning. *IEEE Access*, *7*, 15196–15209. https://doi.org/10.1109/ACCESS.2019.2892066