

**PENGEMBANGAN *DEEP LEARNING* UNTUK MENDETEKSI
SITUS PHISING DENGAN MENGGUNAKAN *CONVOLUTIONAL
NEURAL NETWORK***

TUGAS AKHIR

Alvin Chandra Hadiyanto

1318013



INSTITUT
TEKNOLOGI
HARAPAN
BANGSA
Veritas vos liberabit

**PROGRAM STUDI TEKNIK KOMPUTER
INSTITUT TEKNOLOGI HARAPAN BANGSA
BANDUNG
2022**

**PENGEMBANGAN DEEP LEARNING UNTUK MENDETEKSI
SITUS PHISING DENGAN MENGGUNAKAN CONVOLUTIONAL
NEURAL NETWORK**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh
gelar sarjana dalam bidang Teknik Komputer**

Alvin Chandra Hadiyanto

1318013



INSTITUT
TEKNOLOGI
HARAPAN
BANGSA

Veritas vos liberabit

**PROGRAM STUDI TEKNIK KOMPUTER
INSTITUT TEKNOLOGI HARAPAN BANGSA
BANDUNG**

2022

ABSTRAK

Nama : Alvin Chandra Hadiyanto
Program Studi : Teknik Komputer
Judul : Pengembangan *Deep Learning* Untuk Mendeteksi Situs *Phising* Dengan Menggunakan *Convolutional Neural Network*

Saat ini setiap orang sangat bergantung pada internet. Semua orang melakukan belanja *online* dan *online* aktivitas seperti Bank *online*, pemesanan *online*, isi ulang *online* dan lainnya di internet, namun ada *phising* yang bisa mengambil data pengguna internet. *Phishing* adalah jenis ancaman situs web dengan cara meniru situs web asli secara ilegal untuk mengambil informasi seperti login, id, password dan informasi kartu kredit. Siapapun bisa terkena *phising* karena pengguna kurang mengetahui atau memperhatikan URL dan penyerang juga dapat mempertimbangkan kepribadian dari pengguna. Sistem deteksi *phishing* umumnya dibagi menjadi dua grup: Sistem deteksi berbasis daftar dan berbasis *machine learning*. Dalam *machine learning* dilakukan pengumpulan data kemudian melakukan klasifikasi. Pada penelitian ini dikembangkan pendeteksi situs menggunakan *convolutional neural network* untuk melakukan klasifikasi apakah situs tersebut *phishing* atau *legitimate* dengan arsitektur 3 layer 1D CNN (*Convolutional Neural Network*), 1 layer maxpooling1d, 1 layer flatten, dan 2 layer dense. Hasil dari penelitian ini berhasil mendapat *accuracy* sebesar 0,877101 atau 87,7%. Pada penelitian ini juga dibuat user interface agar pengguna dapat melakukan input sebuah URL (*Uniform Resource Locator*).

Kata kunci: Cybersecurity, Phising, Website, Machine Learning, URL.

ABSTRACT

Name : Alvin Chandra Hadiyanto
Department : Computer Engineering
Title : Pengembangan *Deep Learning* Untuk Mendeteksi Situs *Phising* Dengan Menggunakan *Convolutional Neural Network*

Nowadays everyone is very dependent on the internet. Everyone does online shopping and online activities such as online banking, online ordering, online top-up and others on the internet, but there are phishers that can retrieve internet user data. Phishing is a type of website threat by illegally impersonating a real website to retrieve information such as login, id, password and credit card information. Anyone can get phished because the user doesn't know or pay attention to the URL and the attacker can also consider the personality of the user. Phishing detection systems are generally divided into two groups: List-based and machine learning-based detection systems. In machine learning, data is collected and then classified. In this study, a detection site was developed using a convolutional neural network to classify whether the site is phishing or legitimate with the architecture 3 layer 1D CNN (Convolutional Neural Network), a maxpooling1d layer, a flatten layer, and 2 density layers. The results of this study managed to get a accuracy of 0,877101 or 87,7%. In this study also created a user interface so that users can input a URL (Uniform Resource Locator).

Keywords: Cybersecurity, Phising, Website, Machine Learning, URL.

KATA PENGANTAR

Puji syukur kami panjatkan kehadiran Tuhan, yang telah melimpahkan rahmatNya, sehingga penulis dapat menyelesaikan penyusunan tugas akhir ini yang penulis beri judul ‘Pengembangan *Deep Learning* untuk Mendeteksi Situs *Phising* dengan Menggunakan *Convolutional Neural Network*’.

Karena itu penulis ingin mengucapkan terimakasih kepada pihak-pihak yang telah membantu menyelesaikan penyusunan tugas akhir ini.

1. Tuhan Yang Maha Esa. Atas berkat, anugerah, keajaiban, mukjizat, skenario, rancangan, ide, semangat, kesehatan dan segala yang telah diatur olehNya untuk penulis sehingga dapat menyelesaikan tugas akhir ini hingga mencapai akhir yang melegakan.
2. Orang tua dan keluarga saya yang selalu memberikan dukungan baik secara moril maupun material kepada saya selama penulis menempuh pendidikan sarjana di Institut Teknologi Harapan Bangsa, terutama dalam melakukan penelitian ini.
3. Kedua pembimbing tugas akhir, Bapak Yoyok Gamaliel, M.Eng dan Bapak Maclaurin Hutagalung, Ph.D yang telah berjasa mendukung, membimbing, memberi masukan, arahan, semangat dan motivasi dalam menyelesaikan tugas akhir ini.
4. Kedua penguji tugas akhir, Bapak Dr. Herry I. Sitepu dan Ibu Dina Angela, M.T yang telah berjasa mendukung, membimbing, memberikan masukan, arahan dan semangat dalam menyelesaikan tugas akhir ini.
5. Seluruh staf, dosen, dan keluarga besar Institut Teknologi Harapan Bangsa yang telah memberikan bantuan, fasilitas, dan pelayanan yang baik sehingga proses pengerjaan Tugas Akhir dapat berjalan dengan lancar dan baik.

Juga kepada semua pihak yang telah banyak membantu penyusunan tugas akhir ini, yang tidak dapat disebutkan, penulis mengucapkan terimakasih yang terdalam. Semoga tugas akhir ini dapat menjadi bermanfaat bagi kepentingan kita bersama.

Bandung, 25 Juli 2022

Hormat penulis,

Alvin

DAFTAR ISI

ABSTRAK	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR.....	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN	xi
BAB 1 PENDAHULUAN.....	1-1
1.1 Latar Belakang	1-1
1.2 Rumusan Masalah	1-2
1.3 Tujuan Penelitian	1-2
1.4 Batasan Masalah	1-2
1.5 Hasil dan Manfaat	1-2
1.6 Metodologi Penelitian.....	1-3
BAB 2 KAJIAN REFERENSI	2-1
2.1 Identifikasi Masalah.....	2-1
2.2 Pemangku Kepentingan (<i>StakeHolder</i>).....	2-1
2.3 Riset/ Produk Terkait	2-1
2.4 Sistem yang Sudah Ada	2-2
2.5 Sistem yang Diusulkan.....	2-3
2.6 Analisis Kebutuhan	2-7
BAB 3 PERANCANGAN DAN IMPLEMENTASI	3-1
3.1 Arsitektur Sistem.....	3-1
3.1.1 Dataset.....	3-2
3.1.2 Training	3-5
3.1.3 Model <i>Convolutional Neural Network</i>	3-7
3.1.4 Penilaian Training Model <i>Convolutional Neural Network</i>	3-13
3.1.5 Implementasi Sistem	3-14
BAB 4 PENGUJIAN DAN ANALISIS	4-1
4.1 Metode Pengujian	4-1
4.2 Skenario Pengujian	4-1
4.3 Hasil <i>Learning Curve Loss</i> dan <i>Accuracy</i>	4-1

4.4 Hasil Pengujian Pengaruh Rasio Dataset dan <i>Epoch</i> Terhadap <i>Accuracy</i>	4-5
4.5 Hasil Pengujian Pada Implementasi Sistem.....	4-8
4.6 Pengujian Analisis Kebutuhan	4-12
4.7 Analisis Hasil Pengujian	4-13
BAB 5 KESIMPULAN DAN SARAN	5-1
5.1 Kesimpulan	5-1
5.2 Saran.....	5-1

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	2-6
Tabel 2.2 Kebutuhan Fungsional.....	2-7
Tabel 2.3 Kebutuhan Nonfungsional.....	2-7
Tabel 3.1 Penjelasan Dataset Setelah mengalami Pengurangan <i>Feature</i>	3-3
Tabel 3.2 Pemetaan Tahapan Proses Arsitektur Model.....	3-12
Tabel 3.3 Contoh <i>Confusion Matrix</i>	3-13
Tabel 4.1 Skenario Pengujian.....	4-1
Tabel 4.2 Hasil Pengujian Nilai Akurasi dengan Rasio Dataset 60:20:20 dan <i>Epoch</i>	4-5
Tabel 4.3 Hasil Pengujian Nilai Akurasi dengan Rasio Dataset 70:15:15 dan <i>Epoch</i>	4-6
Tabel 4.4 Hasil Pengujian Nilai Akurasi dengan Rasio Dataset 80:10:10 dan <i>Epoch</i>	4-6
Tabel 4.5 Skenario Pengujian Dataset dan Epoch Terhadap Akurasi.....	4-7
Tabel 4.7 <i>Confusion matrix</i> hasil pengujian dengan URL.....	4-8
Tabel 4.8 Pengujian Terhadap <i>User Interface</i>	4-12
Tabel 4.9 Skenario Pengujian <i>Feature Extraction</i> , URL Berbagai <i>Website</i> dan <i>User Interface</i>	4-12
Tabel 4.10 Hasil Pengujian Fungsional dari Sistem.....	4-13
Tabel 4.11 Hasil Pengujian Nonfungsional dari Sistem.....	4-13
Tabel 5.1 Sample Dataset yang Digunakan Untuk <i>Training</i> , <i>Validation</i> dan <i>Testing</i>	A-1

DAFTAR GAMBAR

Gambar 2.1 Sistem <i>Eksisting</i> [2]	2-2
Gambar 2.2 Sistem diusulkan	2-3
Gambar 2.3 Pemetaan 1D <i>Convolutional Layer</i> [9]	2-4
Gambar 3.1 Arsitektur Sistem.....	3-2
Gambar 3.2 Spektrum Koefisien Korelasi (-1 hingga +1) [16]	3-2
Gambar 3.3 <i>Scatter Plot</i> Menunjukkan Korelasi antara Dua Variabel [16].....	3-3
Gambar 3.4 Pemisahan URL ke <i>Sub-Strings</i> [6]	3-3
Gambar 3.5 <i>Flowchart Preprocessing</i>	3-6
Gambar 3.6 Skema <i>Convolution, Max Pooling, Flatten</i> dan <i>Dense</i>	3-11
Gambar 3.7 <i>Dense Layer</i> [17].....	3-12
Gambar 3.8 <i>User Interface</i> Sistem [18].....	3-14
Gambar 4.1 <i>Training and Validation Loss Curves</i> 60:20:20	4-2
Gambar 4.2 <i>Training and Validation Accuracy Curves</i> 60:20:20	4-2
Gambar 4.3 <i>Training and Validation Loss Curves</i> 70:15:15	4-3
Gambar 4.4 <i>Training and Validation Accuracy Curves</i> 70:15:15	4-3
Gambar 4.5 <i>Training and Validation Loss Curves</i> 80:10:10	4-4
Gambar 4.6 <i>Training and Validation Accuracy Curves</i> 80:10:10	4-4
Gambar 4.7 Pengujian <i>User Interface Legitimate</i>	4-9
Gambar 4.8 Pengujian <i>User Interface Legitimate</i>	4-10
Gambar 4.9 Pengujian <i>User Interface Phishing</i>	4-11

DAFTAR LAMPIRAN

LAMPIRAN A SAMPLE DATASET UNTUK DETEKSI SITUS *WEB PHISHING*A-1

DAFTAR REFERENSI

- [1] O. K. Sahingoz, E. Buber, O. Demir and B. Diri, "Machine Learning Based Phishing Detection from URLs," *Expert Systems With Applications*, vol. 117, pp. 345-357, 2019.
- [2] T. J. Rashid, Mahmood, M. W. Nisar and T. Nazir, "Phishing Detection Using Machine Learning Techniques," in *Smart Systems and Emerging Technologies (SMARTTECH)*, Canada, 2020.
- [3] Avira Operations GmbH & Co. KG, "Avira," Avira Operations GmbH & Co. KG, 1986. [Daring]. Tersedia: <https://www.avira.com/>. [Diakses 11 Oktober 2021].
- [4] eset, "eset," ESET, spol. s r.o., 26 Agustus 1992. [Daring]. Tersedia: <https://www.eset.com/int/>. [Diakses 11 Oktober 2021].
- [5] C. Yao, Y. Qu and B. Jin, "A Convolutional Neural Network Model for Online Medical Guidance," *Institute of Electrical and Electronics Engineers*, vol. 4, pp. 4094 - 4103, 2016.
- [6] G. Vrbančič, I. Fister and V. Podgorelec, "Datasets for phishing websites detection," *Elsevier*, vol. 33, 4 September 2020.
- [7] Minitab, "A comparison of the Pearson and Spearman correlation methods," Minitab Ltd, 2022. [Daring]. Tersedia: <https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/a-comparison-of-the-pearson-and-spearman-correlation-methods/>. [Diakses 2022].
- [8] stackoverflow, "Removing duplicate rows from a csv file using a python script," stackoverflow, April 2013. [Daring]. Tersedia: <https://stackoverflow.com/questions/15741564/removing-duplicate-rows-from-a-csv-file-using-a-python-script>. [Diakses 30 Mei 2022].
- [9] A. Shenfield and . M. Howarth, "A Novel Deep Learning Model for the Detection and Identification of Rolling Element-Bearing Faults," *Sensors*, 2020.
- [10] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Elsevier*, vol. 151, 2020.
- [11] R. S. Srinivasamurthy, "Understanding 1D Convolutional Neural Networks Using Multiclass Time-Varying Signals," *Tigerprints*, vol. 2911, 2018.
- [12] G. V. Situmeang, "Pengenalan Jenis Sampah Menggunakan Metode Convolutional Neural Network," *Final Project ITHB 2021*, 2021.
- [13] Google, "Logistic Regression: Calculating a Probability," Machine Learning Crash

DAFTAR REFERENSI

- Course, 16 6 2021. [Daring]. Tersedia: <https://developers.google.com/machine-learning/crash-course/logistic-regression/calculating-a-probability>. [Diakses 4 June 2022].
- [14] R. Arthana, "Mengenal Accuracy, Precision, Recall dan Specificity serta yang diprioritaskan dalam Machine Learning," medium.com, 5 April 2019. [Daring]. Tersedia: <https://rey1024.medium.com/mengenal-accuracy-precision-recall-dan-specificity-serta-yang-diprioritaskan-b79ff4d77de8>. [Diakses 2 Juni 2022].
- [15] H. N. B, "Confusion Matrix, Accuracy, Precision, Recall, F1 Score," medium.com, 11 Desember 2019. [Daring]. Tersedia: <https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd>. [Diakses 25 Juni 2022].
- [16] N. Gogtay and U. Thatte, "Principles of Correlation Analysis," *Journal of The Association of Physicians of India*, vol. 65, pp. 78-81, 2017.
- [17] P. Sharma, "Keras Dense Layer Explained for Beginners," Making Ai Simple, 20 Oktober 2020. [Daring]. Tersedia: <https://machinelearningknowledge.ai/keras-dense-layer-explained-for-beginners/>. [Diakses 18 Juni 2022].
- [18] nicepage.com, "Contact Form On Gradient HTML Template," Nicepage Free Website Builder Software, 2022. [Daring]. Tersedia: <https://nicepage.com/ht/1975579/contact-form-on-gradient-html-template#>. [Diakses 27 Mei 2022].
- [19] P. Baheti, "Train, Validation, and Test Sets: How to Split Your Machine Learning Data," v7labs, 26 Mei 2022. [Daring]. Tersedia: <https://www.v7labs.com/blog/train-validation-test-set>. [Diakses 14 Juni 2022].
- [20] Z. Wang and Z. Qu, "Research on Web Text Classification Algorithm Based on Improved CNN and SVM," in *Institute of Electrical and Electronics Engineers*, Chengdu, China, 2017.
- [21] Eugen, "What is a Learning Curve in Machine Learning?," Baeldung, 23 Oktober 2020. [Daring]. Tersedia: <https://www.baeldung.com/cs/learning-curve-ml>. [Diakses 27 Juli 2022].
- [22] D. Ulevitch, "PhishTank," Cisco Talos Intelligence Group, Oktober 2006. [Daring]. Tersedia: <https://phishtank.org/>. [Diakses Januari 2022].
- [23] Moz, "The Moz Top 500 Websites," Moz, Inc, 2022. [Daring]. Tersedia: <https://moz.com/top500>. [Diakses 4 Juli 2022].