

**IMPLEMENTASI ALGORITMA *GRADIENT*
BOOSTING UNTUK MENDETEKSI *LINK PHISHING*
BERBASIS *WEBSITE***

SKRIPSI

Oleh:

Febrina Diponegoro

201910225023



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2023**

LEMBAR PERSETUJUAN PEMBIMBING

Judul Skripsi : Implementasi Algoritma *Gradient Boosting*
Untuk Mendeteksi Link Phising Berbasis
Website
Nama Mahasiswa : Febrina Diponegoro
Nomor Pokok Mahasiswa : 201910225023
Program Studi /Fakultas : Informatika/Illmu Komputer
Tanggal Lulus Ujian : 31 Juli 2023

Bekasi, 7 Agustus 2023

MENYETUJUI,

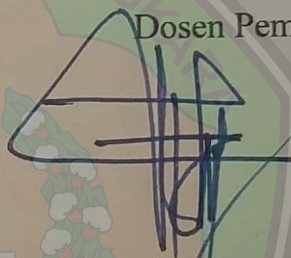
Dosen Pembimbing I



Sugiyatno, S.Kom., M.Kom

NIDN. 0313077206

Dosen Pembimbing II



Ahmad Fathurrozi, S.E., M.M.S.I

NIDN. 0327117402

LEMBAR PENGESAHAN

Judul Skripsi : Implementasi Algoritma *Gradient Boosting* Untuk Mendeteksi Link Phising Berbasis Website

Nama Mahasiswa : Febrina Diponegoro

Nomor Pokok Mahasiswa : 201910225023

Program Studi /Fakultas : Informatika/Illmu Komputer

Tanggal Lulus Ujian : 31 Juli 2023

Bekasi, 7 Agustus 2023

MENGESAHKAN,

Ketua Tim Penguji : Joni Warta, S.Si., M.Si
NIDN. 0317066202

Penguji 1 : Allan Desi Alexander, ST, M.Kom
NIDN. 0305127404

Penguji 2 : Sugiyatno, S.Kom., M.Kom
NIDN. 0313077206

MENGETAHUI,

Ketua
Program Studi Informatika

Ahmad Fathurrozi, S.E., M.M.S.I
NIP. 2012486

Dekan
Fakultas Ilmu Komputer

Dr. Dra. Tyastuti Sri Lestari, MM.
NIP. 1408206

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : FEBRINA DIPONEGORO
NPM : 201910225023
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : IMPLEMENTASI ALGORITMA GRADIENT
BOOSTING UNTUK MENDETEKSI LINK
PHISHING BERBASIS WEBSITE

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Bekasi, 7 Agustus 2023

Penulis



ABSTRAK

Febrina Diponegoro. 201910225023. implementasi algoritma gradient boosting untuk mendeteksi link phishing berbasis website. Universitas Bhayangkara Jakarta Raya. 2023.

Terdapat kerugian yang sangat tinggi dikalangan masyarakat maupun perusahaan yang merupakan dampak dari kejahatan *phishing*. Hal ini terjadi karena tidak adanya pengetahuan serta rasa kewaspadaan masyarakat atau pegawai perusahaan khususnya bank untuk menganalisa, menghindari dan mengidentifikasi *link-link* anomali yang mereka dapat dari media informasi manapun. Pada penelitian ini, dilakukan implementasi algoritma *Gradient Boosting* untuk mendeteksi *link phishing* berbasis *website*. *Phishing* merupakan salah satu metode serangan siber yang berbahaya, di mana penyerang mencoba untuk menipu pengguna dengan menyajikan tautan palsu yang meniru situs *web* resmi untuk mencuri informasi sensitif pengguna. Tujuan dari penelitian ini adalah untuk mengembangkan sebuah aplikasi *web* menggunakan *Flask*, sebuah *framework web Python*, untuk memeriksa apakah sebuah *URL* aman atau tidak aman berdasarkan ekstraksi fitur dari link tersebut. Penelitian ini memberikan kontribusi dalam pengembangan solusi deteksi *phishing* yang efektif dan dapat diimplementasikan dalam aplikasi *web*. Model yang dihasilkan dari implementasi algoritma *Gradient Boosting* dapat digunakan sebagai salah satu langkah pencegahan dalam melindungi pengguna dari ancaman keamanan siber yang serius dan melindungi informasi sensitif dari pencurian oleh pihak yang tidak bertanggung jawab.

Kata kunci: *Phishing, Website, Gradient BoostingBank*

ABSTRACT

Febrina Diponegoro. 201910225023. *implementation of gradient boosting algorithm to detect website-based phishing links. Jakarta Bhayangkara University. 2023.*

There are very high losses among the public and companies which are the impact of phishing crimes. This happens due to the lack of knowledge and a sense of alertness by the public or company employees, especially banks, to analyze, avoid and identify anomalous links that they get from any media information. In this study, the implementation of the Gradient Boosting algorithm was carried out to detect phishing web-based links. Phishing is one of the most dangerous cyberattack methods, in which attackers try to deceive users by serving fake links that mimic legitimate websites to steal sensitive user information. The aim of this research is to develop a web application using Flask, a Python web framework, to check whether a URL is safe or insecure based on feature extraction from the link. This research contributes to the development of effective phishing detection solutions that can be implemented in web applications. The model resulting from the implementation of the Gradient Boosting algorithm can be used as a preventive measure in protecting users from serious cybersecurity threats and protecting sensitive information from theft by irresponsible parties.

Keywords: *Phishing, Websites, Gradient Boosting, Banks*

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini:

Nama : Febrina Diponegoro
NPM : 201910225023
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalty-Free Right)**, atas karya ilmiah saya yang berjudul:
Implementasi Algoritma Gradient Boosting untuk Mendeteksi Link Phishing Berbasis Website

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bekasi
Pada tanggal : 8 Agustus 2023
Yang Menyatakan



Febrina Diponegoro

KATA PENGANTAR

Segala puji bagi Allah SWT yang telah melimpahkan banyak nikmat sehingga penulis dapat menyelesaikan penulisan yang berjudul “Implementasi Algoritma Gradient Boosting untuk Mendeteksi Link Phishing Berbasis Website” secara maksimal dan optimal. Shalawat serta salam senantiasa tersampaikan kepada junjungan kita Nabi Muhammad SAW yang telah begitu banyak mengajarkan kebijakan dan menyebarkan ilmunya pada semua umatnya.

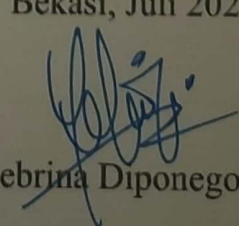
Terselesaikannya penulisan ini, berkat dukungan, bantuan, dorongan, petunjuk serta saran dari berbagai pihak oleh karena itu, penulis ingin menyampaikan ucapan terima kasih khusus kepada orang tua yang penulis sayangi, Bapak Ir. Ridwan Diponegoro dan Ibu Jeni Arisandi, yang telah memberi dukungan selama proses perkuliahan dan penulisan skripsi ini. Penulis juga mengucapkan terima kasih kepada :

1. Bapak Irjen Pol (purn) Prof. Dr. Bambang Karsono, S.H., M.M. selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Ibu Dr. Dra. Tyastuti Sri Lestari, M.M., selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Sugiyatno, S.Kom., M.Kom selaku Dosen Pembimbing 1 penyusunan skripsi.
4. Bapak Ahmad Fathurrozi, S.E., M.M.S.I selaku Ketua Program Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya juga selaku Dosen Pembimbing 2 penyusunan skripsi.

5. Bapak Prio Kustanto S.T., M.Kom, selaku Dosen Pembimbing akademik yang sangat ramah dan kompeten dalam melaksanakan tugasnya sebagai dosen pembimbing akademik TIFA1.
6. Muhammad Arsalan Diponegoro S.Kom, Muhammad Haikal Diponegoro, selaku kakak dan adik yang sangat memotivasi penulis untuk terus belajar dan berprestasi.
7. GDSCUBJ (Google Developer Student Club) yang sampai saat ini memberikan dukungan kepada penulis.
8. Seluruh teman - teman TIFA1 angkatan 2019 yang sama – sama berjuang saling mendoakan serta memberikan semangat hingga penulisan ini selesai.
9. Arya Batara Sena, Lidya Dwi Astuti, Martaulina, Roberto Hutapea S.Kom., Agung Wahyu Prayogo, Hana Lareina selaku rekan penulis selama perkuliahan yang selalu ada dan memberikan dukungan, arahan, motivasi serta semangat sehingga penulis dapat menyelesaikan perkuliahan dan penulisan.

Penulis menyadari penulisan ini masih jauh dari kata sempurna karena keterbatasan waktu, dan kurangnya pengetahuan dan kemampuan penulis. Oleh karena itu, kritik dan saran yang bersifat membangun sangat diharapkan demi sempurnanya penyusunan skripsi ini. Semoga skripsi ini bermanfaat bagi semua pihak. Aamiin.

Bekasi, Juli 2023



Febrina Diponegoro

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iv
ABSTRAK	v
ABSTRACT	vi
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	vii
KATA PENGANTAR.....	viii
DAFTAR TABEL	xv
DAFTAR GAMBAR.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	5
1.3 Rumusan Masalah	5
1.4 Batasan Masalah.....	5
1.5 Tujuan dan Manfaat Penelitian.....	6
1.5.1 Tujuan Penelitian	6
1.5.2 Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	9
2.1 Penelitian Relevan.....	9
2.2 <i>Cyber Crime</i>	13

2.3	<i>Cyber Security</i>	14
2.4	<i>Cyber Security Awareness</i>	15
2.5	Keamanan Informasi	16
2.5.1	Aspek – Aspek Keamanan Informasi	17
2.6	Pengertian <i>Phishing</i>	18
2.6.1	Cara Kerja <i>Phishing</i>	19
2.7	Pengertian <i>Website</i>	20
2.7.1	Unsur – Unsur <i>Website</i>	20
2.8	Pengertian <i>Ai</i>	24
2.9	<i>Gradient Boosting</i>	24
2.12	<i>Activity Diagram</i>	26
BAB III METODOLOGI PENELITIAN		28
3.1	Objek Penelitian	28
3.1.1	Profil Perusahaan	28
3.2	Kerangka Penelitian	30
3.3	Metode Pengumpulan Data	31
3.3.1	Wawancara	31
3.3.2	Observasi	31
3.3.3	Studi Pustaka	31
3.4	Metode Analisis	32
3.4.1	Analisis Permasalahan	32
3.4.2	Analisis Sistem Usulan	32

3.4.3	Analisis Kebutuhan Sistem	34
3.5	Dataset	35
3.6	<i>Familiarizing data and Exploratory Data Analyst</i>	36
3.7	<i>Features pada Dataset</i>	37
3.3.1	<i>Index</i>	37
3.3.2	Menggunakan IP	37
3.3.3	<i>LongURL</i>	37
3.3.4	<i>ShortURL</i>	37
3.3.5	<i>Symbol@</i>	38
3.3.6	<i>Redirecting//</i>	38
3.3.7	<i>PrefixSuffix-</i>	38
3.3.8	<i>SubDomains</i>	38
3.3.9	<i>HTTPS</i>	38
3.3.10	<i>DomainRegLen</i>	38
3.3.11	<i>Favicon</i>	39
3.3.12	<i>NonStdPort</i>	39
3.3.13	<i>HTTPSDomainURL</i>	39
3.3.14	<i>RequestURL</i>	39
3.3.15	<i>AnchorURL</i>	39
3.3.16	<i>LinkInScriptTags</i>	39
3.3.17	<i>ServerFromHanlder</i>	40
3.3.18	<i>InfoEmail</i>	40
3.3.19	<i>AbnormalURL</i>	40

3.3.20	<i>Website Forwarding</i>	40
3.3.21	<i>StatusBarCust</i>	40
3.3.22	<i>DisableRightClick</i>	41
3.3.23	<i>UsingPopupWindow</i>	41
3.3.24	<i>IframeRedirection</i>	41
3.3.25	<i>AgeofDomain</i>	41
3.3.26	<i>DNSRecording</i>	41
3.3.27	<i>WebsiteTraffic</i>	41
3.3.28	<i>PageRank</i>	42
3.3.29	<i>GoogleIndex</i>	42
3.3.30	<i>LinkpointingToPage</i>	42
3.3.31	<i>StatsReport</i>	42
3.3.32	<i>Class</i>	42
3.8	Memvisualisasikan Data.....	44
3.9	Memisahkan dan Membagi <i>Dataset</i>	48
3.10	<i>Build and Model Training</i>	49
3.10.1	<i>Gradient Boosting</i>	50
3.11	<i>Comparing of Model</i>	52
3.12	Menyimpan Model Terbaik	53
3.13	Mengubah Menjadi Format <i>Pickle</i>	54
3.14	<i>Import Framework</i> dan modul.....	54
3.15	<i>Import Model</i>	55
3.16	<i>app.py</i>	56

3.17	Membuat <i>Integration.py</i>	57
3.18	<i>feature.py</i>	58
BAB IV HASIL DAN PEMBAHASAN		59
4.1	Hasil Penelitian.....	59
4.2	Bentuk <i>Website</i>	59
4.3	Identifikasi <i>Link</i>	60
4.4	<i>Output</i>	61
4.5	Hasil <i>Phishing</i>	62
4.6	<i>Link</i> mengandung <i>malware</i> atau <i>ransomeware</i>	62
BAB V PENUTUP		64
5.1	Kesimpulan.....	64
5.2	Saran.....	64
DAFTAR PUSTAKA		65
LAMPIRAN-LAMPIRAN		67

DAFTAR TABEL

Tabel 2 1 Activity Diagram.....	20
---------------------------------	----



DAFTAR GAMBAR

Gambar 2 1 Ilustrasi	13
Gambar 2 2 Ilustrasi Keamanan Informasi	16
Gambar 2 3 Pilar Keamanan <i>Security</i>	17
Gambar 2 4 ilustrasi <i>phishing</i>	18
Gambar 2 5 Cara Kerja <i>Phishing</i>	20
Gambar 3. 1 Kerangka Penelitian.....	30
Gambar 3. 2 sistem usulan	33
Gambar 3. 3 <i>Dataset</i>	35
Gambar 3. 4 <i>sample</i> dan <i>features</i>	36
Gambar 3. 5 <i>list feature</i>	36
Gambar 3. 6 Info <i>Feature</i>	43
Gambar 3. 7 visualisasi data.....	44
Gambar 3. 8 data <i>feature</i>	45
Gambar 3. 9 <i>pairplot</i> fitur	45
Gambar 3. 10 Grafik	46
Gambar 3. 11 <i>URL Count</i>	46
Gambar 3. 12 <i>Diagram Count</i>	47
Gambar 3. 13 Membagi <i>dataset</i>	48
Gambar 3. 14 <i>Model and Training</i>	49
Gambar 3. 15 <i>Gradient Boosting</i>	50
Gambar 3. 16 <i>Predict gr</i>	50

Gambar 3. 17 <i>Computing gr</i>	51
Gambar 3. 18 <i>Result</i>	51
Gambar 3. 19 <i>Comparing Model</i>	52
Gambar 3. 20 <i>fit best model</i>	53
Gambar 3. 21 <i>pickle formating</i>	54
Gambar 3. 22 <i>Import Framework</i>	54
Gambar 3. 23 <i>Import Model</i>	55
Gambar 3. 24 <i>app.py</i>	56
Gambar 3. 25 <i>integration.py</i>	57
Gambar 3. 26 <i>payload</i>	57
Gambar 3. 27 <i>Scoring</i>	57
Gambar 3. 28 <i>feature.py</i>	58
Gambar 4. 1 halaman <i>website</i>	59
Gambar 4. 2 <i>identification link</i>	60
Gambar 4. 3 <i>output</i>	61
Gambar 4. 4 hasil <i>phishing</i>	62
Gambar 4. 5 link mengandung <i>malware</i> dan <i>ransomeware</i>	62