



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

SKRIPSI

**MUHAMAD ABDUL GHANNI AL GHIFARI
1810511054**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN
JAKARTA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
2022**



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

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
Sarjana Komputer**

**MUHAMAD ABDUL GHANNI AL GHIFARI
1810511054**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN
JAKARTA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
2022**

PERNYATAAN ORIGINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber yang dikutip serta dirujuk telah saya nyatakan dengan benar.

Nama : Muhamad Abdul Ghanni Al Ghifari
NIM : 1810511054
Tanggal : 19 Juli 2022

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 24 Juli 2022

Yang Menyatakan,



(Muhamad Abdul Ghanni Al Ghifari)

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Muhamad Abdul Ghanni Al Ghifari
NIM : 1810511054
Fakultas : Fakultas Ilmu Komputer
Program Studi : S1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non Eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

Implementasi Ekstensi Google Chrome Dalam Mendeteksi Situs Web Phishing Menggunakan Algoritma Random Forest

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Dibuat di : Jakarta

Pada Tanggal : 24 Juli 2022

Yang Menyatakan,



(Muhamad Abdul Ghanni Al Ghifari)

LEMBAR PENGESAHAN

Dengan ini menyatakan bahwa Tugas Akhir berikut:

Nama : Muhamad Abdul Ghanni Al Ghifari
NIM : 1810511054
Program Studi : S1 Informatika
Judul : Implementasi Ekstensi Google Chrome Dalam Mendeteksi Situs Web *Phishing* Menggunakan Algoritma *Random Forest*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Dr. Ermatita, M.Kom.

Penguji I



Desta Sandya Pravista, S.Komp., M.Kom.

Penguji II



Bayu Hananto, S.Kom., M.Kom.

Pembimbing I



Bambang Tri Wahyono, S.Komp., M.Si.

Pembimbing II



Dr. Ermatita, M.Kom.
Dekan



Desta Sandya Prasvita, S.Komp., M.Kom.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Pengesahan : 25 Juli 2022



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.*

KATA PENGANTAR

Puji Syukur saya ucapkan kepada Allah SWT, atas karunia yang telah diberikan, sehingga penulisan Skripsi dapat selesai dikerjakan secara tepat waktu. Penelitian ini disusun sebagai persyaratan untuk menyelesaikan studi Sarjana Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta. Skripsi ini berjudul “Implementasi Ekstensi Google Chrome Dalam Mendeteksi Situs Web *Phishing* Menggunakan Algoritma *Random Forest*”.

Penulis ingin mengungkapkan banyak terima kasih kepada pihak yang telah membantu dalam segi apapun, yang penulis tujukan kepada:

1. Ibu Dr. Ermatita, M.Kom. selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.
2. Bapak Desta Sandya Prasvita, S.Komp., M.Kom. selaku Ketua Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.
3. Bapak Bayu Hananto, S.Kom, M.Kom. selaku Dosen Pembimbing yang telah memberikan saran dan bimbingan yang bermanfaat dari awal pembuatan proposal hingga skripsi.
4. Bapak Bambang Tri Wahyono, S.Kom, M.Si. selaku Dosen Pembimbing yang telah memberikan saran dan bimbingan yang bermanfaat dari awal pembuatan proposal hingga skripsi.
5. Orang tua serta keluarga yang tak pernah berhenti mendoakan serta memberikan dukungan dalam berbagai hal.
6. Teman-teman seperjuangan Program Studi Informatika Angkatan 2018 yang telah memberikan semangat untuk menyelesaikan skripsi ini.

Penulis menyadari masih banyaknya kekurangn dan masih jauh dari kata sempurna dalam menyusun skripsi ini. Akhir kata penulis ucapkan terimakasih.

Jakarta, 25 Juli 2022

DAFTAR ISI

PERNYATAAN ORIGINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Ruang Lingkup	3
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	3
1.6. Luaran yang Diharapkan	3
1.7. Sistematika Penulisan	4
BAB 2 TINJAUAN PUSTAKA	6
2.1. Phishing	6
2.2. Machine Learning	7
2.3. Bagging	9
2.4. Random Forest	10
2.5. Ekstensi Google Chrome	11

2.6.	Penelitian Terdahulu	12
BAB 3 METODOLOGI PENELITIAN		18
3.1.	Gambaran Umum Penelitian	18
3.2.	Studi Literatur.....	19
3.3.	Pengumpulan Data.....	19
3.4.	Pre-processing Data	19
3.5.	Model Klasifikasi Random Forest.....	19
3.6.	Kinerja Model.....	21
3.7.	Ekstensi Chrome dan Implementasi Model	22
3.7.1.	Sistem Arsitektur	23
3.7.2.	Ekstraksi Fitur	24
3.7.3.	Format Model untuk Implementasi	24
3.8.	Analisis dan Evaluasi Hasil Penelitian	24
3.9.	Alat Bantu Penelitian	25
3.10.	Jadwal Rencana Penelitian	25
BAB 4 HASIL DAN PEMBAHASAN		27
4.1.	Data.....	27
4.1.1.	Fitur berbasis URL	27
4.1.2.	Fitur Berbasis Source Code Pada Halaman	28
4.1.3.	Fitur Berbasis Domain	30
4.2.	Pre-processing Data	32
4.3.	Pelatihan Model.....	33
4.4.	Evaluasi Model.....	35
4.5.	Mengubah Model ke Format JSON.....	36
4.6.	Ekstensi Chrome dan Implementasi Model	38
4.7.	Ekstraksi Fitur Pada Situs Web	38

4.8.	Klasifikasi Situs Web.....	39
4.9.	Tampilan Ekstensi Chrome	40
4.10.	Hasil Dari Kinerja Ekstensi Chrome	40
BAB 5 PENUTUP		42
5.1.	Kesimpulan.....	42
5.2.	Saran	43
DAFTAR PUSTAKA		44
RIWAYAT HIDUP		46
LAMPIRAN		47

DAFTAR GAMBAR

Gambar 2. 1 Teknik Random Forest.....	10
Gambar 2. 2 Komponen dan Bagaimana Suatu Ekstensi Bekerja.....	11
Gambar 3. 1 Diagram Kerangka Penelitian	18
Gambar 3. 2 Deskripsi Tahapan Dari Pembuatan Model Random Forest.....	20
Gambar 3. 3 Diagram Sistem Arsitektur Ekstensi Peramban	23
Gambar 4. 1 Grafik Bar dari Feature Important	34
Gambar 4. 2 Visualisasi Pohon pada Salah Satu Decision Tree	35
Gambar 4. 3 Pseudocode Mengubah Model ke JSON.....	37
Gambar 4. 4 Bentuk Sebagian Isi Model Setelah Diubah ke JSON	37
Gambar 4. 5 Tampilan Popup Ekstensi.....	40
Gambar 4. 6 Tampilan bila mengunjungi situs phishing	40

DAFTAR TABEL

Tabel 2. 1 Contoh Suatu Dataset	9
Tabel 2. 2 Review Penelitian Terdahulu	12
Tabel 3. 1 Confusion Matrix untuk Dua Kelas.....	22
Tabel 3. 2 Tabel Jadwal Penelitian	25
Tabel 4. 1 Dataset yang Digunakan	31
Tabel 4. 2 Fitur yang Digunakan Setelah Seleksi Fitur	32
Tabel 4. 3 Jumlah data setelah pembagian data	33
Tabel 4. 4 Tabel Confusion Matrix.....	35
Tabel 4. 5 Nilai fitur-fitur setelah diekstrak pada suatu situs web	39
Tabel 4. 8 Tabel Confusion matrix	41

DAFTAR LAMPIRAN

Lampiran 1 – Script Python untuk membuat model.....	47
Lampiran 2 – Manifest Ekstensi Browser	49
Lampiran 3 – HTML yang Ada pada Ekstensi Browser.....	50
Lampiran 4 – JavaScript Ekstensi Browser.....	51
Lampiran 5 - Tabel data yang diambil dari Phistank.org dan RankRanger.com ..	59
Lampiran 6 – Hasil Turnitin	63