



IMPLEMENTASI ALGORITMA KLASIFIKASI *RANDOM FOREST* DAN *KNN* UNTUK MENDETEKSI WEBSITE *PHISHING*

LAPORAN SKRIPSI

REIHAN SETYA ABIDA

41519010177

UNIVERSITAS

MERCU BUANA

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA JAKARTA
2023**

HALAMAN PENYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Reihan Setya Abida
NIM : 41519010177
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Implementasi Algoritma Klasifikasi Random Forest Dan Knn Untuk Mendeteksi Website Phishing

Menyatakan bahwa Laporan Skripsi ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.



UNIVERSITAS
MERCU BUANA

Jakarta, 3 Juli 2023



Reihan Setya Abida.

HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama : Reihan Setya Abida

NIM : 41519010177

Program Studi : Teknik Informatika

Judul Laporan Skripsi : Implementasi Algoritma Klasifikasi Random Forest Dan Knn Untuk Mendeteksi Website Phishing

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Lukman Hakim, ST,
M.Kom

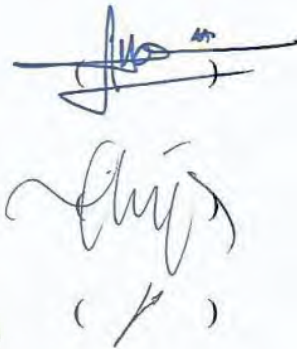
NIDN : 0327107701

Ketua Penguji : Dr. Ir. Eliyani

NIDN : 0321026901

Penguji 1 : Dr. Leonard Goeirmanto

NIDN : U 0312087601



MERCU BUANA

Jakarta, 1 Agustus 2023

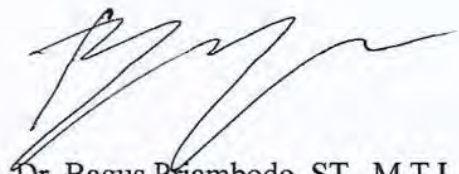
Mengetahui,

Dekan

Ketua Program Studi



Dr. Bambang Jekonowo, S.Si., M.T.I



Dr. Bagus Priambodo, ST., M.T.I

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Laporan Skripsi ini. Penulisan Laporan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Prof.Dr.andi Adriansyah, M.Eng selaku Rektor Universitas Mercu Buana.
2. Yaya Sudarya Triyana,Ph.D selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Bagus Priambodo, ST, MTI selaku Ketua Program Studi Teknik Informatika.
4. Bapak Lukman Hakim, ST, M.Kom, selaku Dosen Pembimbing saya telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan skripsi ini.
5. Bapak Wawan Gunawan, S.Kom, MT selaku Dosen Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.
6. Orang tua saya yang telah mensupport saya dan memberikan semangat kepada saya sehingga saya dapat menyelesaikan tugas akhir saya.
7. Teman-teman dan pasangan saya Lisa Maulina yang telah membantu dan menjadi support system selama proses tugas akhir saya.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Skripsi ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 3 Juli 2023

Reihan Setya Abida

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Reihan Setya Abida
NIM : 41519010177
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Implementasi Algoritma Random Forest dan KNN Untuk Mendeteksi Website Phishing

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif** (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 3 Juli 2023

Yang menyatakan,



(Reihan Setya Abida)

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERNYATAAN KARYA SENDIRI.....	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS.....	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	2
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah.....	3
BAB II	4
TINJAUAN PUSTAKA	4
2.1 Penelitian Terdahulu	4
2.2 Teori Pendukung	8
2.2.1 Website Phising.....	8
2.2.2 Data Mining	8
2.2.2 Klasifikasi	8
2.2.4 Random Forest	9
2.2.5 KNN	10
BAB III.....	12
METODE PENELITIAN	12

3.1 Jenis Penelitian.....	12
3.2 Metode Pengumpulan Data	12
3.3 Tahapan Penelitian	12
3.3.1 Dataset.....	13
3.3.2 Preprocessing Data	14
3.3.3 Pengujian Algoritma	14
3.3.4 Uji Validasi	14
3.3.5 Analisis dan Hasil	14
BAB IV	15
HASIL PEMBAHASAN.....	15
4.1 Dataset.....	15
4.2 Pre-Processing.....	17
4.3 Pembuatan Model.....	25
4.4 Visualisi Data	27
4.5 Pengujian.....	29
4.6 Analisis Hasil	31
BAB V.....	35
KESIMPULAN DAN SARAN	35
5.1 Kesimpulan	35
5.2 Saran.....	35
DAFTAR PUSTAKA.....	36
LAMPIRAN.....	38

DAFTAR TABEL

Tabel 1 Penelitian Terkait	4
----------------------------------	---



DAFTAR GAMBAR

Gambar 1. Grafik target phishsing	2
Gambar 2. Tahapan Metodologi Penelitian	13
Gambar 3 Dataset malicious-phish	15
Gambar 4 info dataset	16
Gambar 5 count value null	16
Gambar 6 cek value null	16
Gambar 7 value atribut type	16
Gambar 8 dataset atribut type	17
Gambar 9 balance dataset dengan downsampling dan upsampling	18
Gambar 10 dataset setelah menghapus www	19
Gambar 11 Mengubah atribut type	19
Gambar 12 parameter domain	20
Gambar 13 parameter abnormal_url	20
Gambar 14 atribut https	21
Gambar 15 atribut ip address	21
Gambar 16 atribut symbol@	22
Gambar 17 atribut symbol?	22
Gambar 18 atribut symbol%	22
Gambar 19 atribut url_length	23
Gambar 20 atribut shortUrl	23
Gambar 21 atribut digits	24
Gambar 22 atribut safe_url	24
Gambar 23 total atribut	25
Gambar 24 Pembuatan model dataset	26
Gambar 25 model y_train	26
Gambar 26 model X_train	26
Gambar 27 visualisasi barplot atribut type	27
Gambar 28 visualisasi heatmap semua atribut	28
Gambar 29 Hasil algoritma KNN	29
Gambar 30 Hasil algoritma Random Forest	29
Gambar 32 <i>Confusion matrix Random Forest</i>	30
Gambar 31 Confusion matrix KNN	30
Gambar 33 feature importance Random Forest	31
Gambar 34 Perbandingan algoritma	32
Gambar 35 data url phishing dan safe	33
Gambar 36 test data	33
Gambar 37 uji coba url phishing	34
Gambar 38 uji coba url safe	34

DAFTAR LAMPIRAN

lampiran 1. Bimbingan.....	38
lampiran 2. Luaran Tugas Akhir	39
lampiran 3. Bukti Submit Jurnal	40
lampiran 4. Naskah Artikel Jurnal	41
lampiran 5. Curriculum Vitae	49
lampiran 6. SP Hak Cipta.....	51
lampiran 7. SPH Hak Cipta.....	52
lampiran 8. Scan KTP	53

