

**IMPLEMENTASI RECURSIVE FEATURE ELIMINATION DAN  
SUPPORT VECTOR MACHINE UNTUK DETEKSI WEBSITE PHISHING**



**SKRIPSI**

**Feliciano Surya Marcello  
00000043255**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG  
2024**

**IMPLEMENTASI RECURSIVE FEATURE ELIMINATION DAN  
SUPPORT VECTOR MACHINE UNTUK DETEKSI WEBSITE PHISHING**



**Feliciano Surya Marcello**

**00000043255**

**UMMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2024**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Feliciano Surya Marcello

NIM : 00000043255

Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Skripsi saya yang berjudul:  
**Implementasi Recursive Feature Elimination dan Support Vector Machine  
Untuk Deteksi Website Phising**

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 8 Mei 2024



(Feliciano Surya Marcello)

## HALAMAN PENGESAHAN

Skripsi dengan judul

### IMPLEMENTASI RECURSIVE FEATURE ELIMINATION DAN SUPPORT VECTOR MACHINE UNTUK DETEKSI WEBSITE PHISHING

oleh

Nama : Feliciano Surya Marcello  
NIM : 00000043255  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Senin, 27 Mei 2024

Pukul 15.00 s/s 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang



(Dr. Maria Irmina Prasetyowati, S.Kom.,  
M.T.)

NIDN: 0725057201

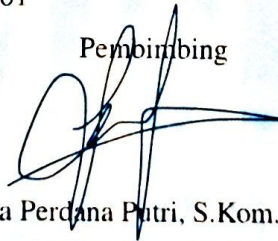
Penguji



(Dennis Gunawan, S.Kom., M.Sc.)

NIDN: 0320059001

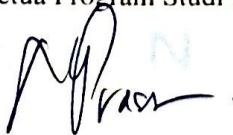
Pembimbing



(Farica Perdana Putri, S.Kom., M.Sc.)

NIDN: 0331019301

PJS Ketua Program Studi Informatika,



(Dr. Eng. Niki Prastomo, S.T., M.Sc.)

NIDN: 0419128203

**LEMBAR PERSETUJUAN PUBLIKASI  
KARYA ILMIAH MAHASISWA**

Yang bertanda tangan dibawah ini:

Nama : Feliciano Surya Marcello  
Nomor Induk Mahasiswa : **00000043255**  
Program Studi : Informatika  
Jenjang : S1  
Judul Karya Ilmiah :

**IMPLEMENTASI RECURSIVE FEATURE ELIMINATION DAN  
SUPPORT VECTOR MACHINE UNTUK DETEKSI WEBSITE  
PHISHING**

Menyatakan dengan sesungguhnya bahwa saya bersedia:

Memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.

Saya tidak bersedia, dikarenakan:

Dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)\*.

Tangerang, 27 Mei 2024



(Feliciano Surya Marcello)

\* Jika tidak bisa membuktikan LoA jurnal/HKI selama 6 bulan kedepan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

**Halaman Persembahan / Motto**

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa karena atas berkat dan hikmatnya, laporan skripsi yang berjudul Implementasi *Recursive Feature Elimination* dan *Support Vector Machine* untuk deteksi *website phishing* dapat terselesaikan dengan baik. Laporan skripsi ini dibuat sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom,) di fakultas Teknik dan Informatika jurusan Informatika Universitas Multimedia Nusantara Tangerang.

Penulis berterima kasih kepada:

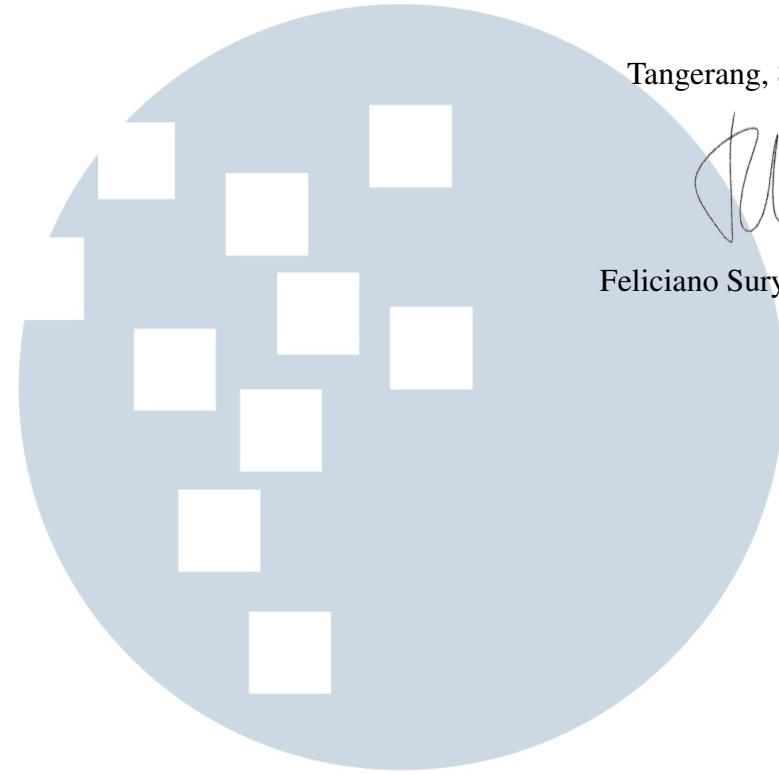
1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Pjs. Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Farica Perdana Putri, S.Kom., M.Sc., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua dan saudara-saudara saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan skripsi ini.
6. Teman-teman Ministry House yang sudah memberikan bantuan dukungan moral, sehingga penulis mendapatkan semangat dalam menyelesaikan skripsi ini.
7. Teman-teman Turquoise yang memberikan dukungan dan masukan, sehingga penulis dapat menyelesaikan skripsi.
8. Teman-teman Organic Ministry yang memberikan dukungan moral sehingga penulis dapat menyelesaikan skripsi.
9. Teman-teman Informatika yang telah memberikan dukungan moral sehingga penulis dapat menyelesaikan skripsi

Semoga laporan skripsi ini bermanfaat sebagai sumber informasi dan inspirasi bagi para pembaca.

Tangerang, 8 Mei 2024



Feliciano Surya Marcello



UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



# IMPLEMENTASI RECURSIVE FEATURE ELIMINATION DAN SUPPORT VECTOR MACHINE UNTUK DETEKSI WEBSITE PHISHING

Feliciano Surya Marcello

## ABSTRAK

Perkembangan teknologi semakin hari semakin pesat terutama pada bidang Teknologi Informasi. *Website* menjadi salah satu akses informasi yang banyak digunakan oleh manusia untuk mendapatkan informasi, melakukan proses bisnis, dan lain-lain. Perkembangan dan kegunaan *website* yang semakin canggih juga membuat banyak pihak yang tidak bertanggung jawab melakukan kejahatan untuk mendapatkan keuntungan. *Website phishing* adalah cara yang populer untuk melakukan penipuan melalui teknologi informasi. Fitur-fitur yang ada dalam *website* menjadi salah satu sarana yang digunakan untuk melakukan *phishing*. Algoritma Support Vector Machine (SVM) adalah salah satu cara yang dapat diimplementasikan dalam mendeteksi *website phishing* dengan melakukan klasifikasi melalui pengecekan fitur-fitur *website*. Namun, algoritma SVM tidak mampu mendeteksi fitur yang banyak sehingga tingkat akurasi dan optimasi yang dihasilkan juga tidak baik. Berdasarkan dataset yang pernah dicoba, algoritma SVM hanya mendapatkan sekitar 60% sampai 70% akurasi. Penggunaan *feature selection Recursive Feature Elimination* (RFE) adalah salah satu cara yang dapat dilakukan untuk menutupi kekurangan dari SVM. Dengan melakukan eliminasi fitur yang tidak berpengaruh, RFE membuat algoritma SVM mendapatkan tingkat akurasi yang lebih tinggi pada dataset yang tersedia yaitu peningkatan sebesar 0.15% dengan akurasi sebesar 96.09%.

**Kata kunci:** *Klasifikasi, Feature Selection, Recursive Feature Elimination, Support Vector Machine, Website Phishing Detection*

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

# **Implementation of Recursive Feature Elimination and Support Vector Machine for Phishing Website Detection.**

Feliciano Surya Marcello

## **ABSTRACT**

*The rapid advancement of technology, especially in the field of Information Technology, is undeniable. Websites have become one of the primary means through which people access information, conduct business processes, and more. However, the increasing sophistication and utility of websites have also led to the proliferation of malicious activities aimed at exploiting them for profit. Phishing websites are a popular method used by cybercriminals to deceive individuals through information technology. Features present in websites are often exploited as means to carry out phishing attacks. Support Vector Machine (SVM) algorithm is one approach that can be implemented to detect phishing websites by classifying website features. However, SVM algorithms may struggle to detect a wide range of features, leading to suboptimal accuracy and performance. Based on the datasets experimented with, SVM algorithms typically achieve only around 60% to 70% accuracy. To address this limitation, the use of Recursive Feature Elimination (RFE) can be employed. By eliminating irrelevant features, RFE enhances the accuracy of SVM algorithms. In the dataset used, this results in an increase in accuracy of 0.15%, with an overall accuracy of 96.09%.*

**Keywords:** *Classification, Feature Selection, Recursive Feature Elimination, Support Vector Machine, Website Phishing Detection*

UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .	ii
HALAMAN PENGESAHAN . . . . .	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	iv
HALAMAN PERSEMBAHAN/MOTO . . . . .	v
KATA PENGANTAR . . . . .	vi
ABSTRAK . . . . .	viii
ABSTRACT . . . . .	ix
DAFTAR ISI . . . . .	x
DAFTAR TABEL . . . . .	xii
DAFTAR GAMBAR . . . . .	xiii
Listings . . . . .	xiv
DAFTAR LAMPIRAN . . . . .	xv
<b>BAB 1 PENDAHULUAN . . . . .</b>	<b>1</b>
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	3
1.3 Batasan Permasalahan . . . . .	4
1.4 Tujuan Penelitian . . . . .	4
1.5 Manfaat Penelitian . . . . .	4
1.6 Sistematika Penulisan . . . . .	4
<b>BAB 2 LANDASAN TEORI . . . . .</b>	<b>6</b>
2.1 Website Phishing . . . . .	6
2.2 Recursive Feature Elimination (RFE) . . . . .	6
2.3 Support Vector Machine (SVM) . . . . .	7
2.4 Standardization . . . . .	11
2.5 Confusion Matrix . . . . .	11
<b>BAB 3 METODOLOGI PENELITIAN . . . . .</b>	<b>14</b>
3.1 Tahap Pelaksanaan Penelitian . . . . .	14
3.1.1 Telaah Literatur . . . . .	14
3.1.2 Pengumpulan Data . . . . .	14
3.1.3 Implementasi Algoritma . . . . .	18
3.1.4 Uji coba dan Evaluasi . . . . .	18
3.1.5 Konsultasi dan Penulisan Laporan . . . . .	19
3.2 Perancangan Sistem . . . . .	19
3.2.1 Flowchart Utama . . . . .	19
3.2.2 Flowchart Preprocessing . . . . .	21
3.2.3 Flowchart Recursive Feature Elimination . . . . .	22
3.2.4 Flowchart Support Vector Machine . . . . .	23
<b>BAB 4 HASIL DAN DISKUSI . . . . .</b>	<b>25</b>
4.1 Spesifikasi Perangkat . . . . .	25
4.2 Implementasi Sistem . . . . .	25
4.2.1 Implementasi Data Cleaning . . . . .	26
4.2.2 Implementasi Recursive Feature Elimination (RFE) . . . . .	28
4.2.3 Implementasi Evaluasi Recursive Feature Elimination . . . . .	29
4.2.4 Implementasi GridSearchCV . . . . .	29
4.2.5 Implementasi Support Vector Machine . . . . .	30
4.2.6 Implementasi Confusion Matrix . . . . .	30
4.3 Uji Coba Sistem dan Hasil Analisis . . . . .	31

4.3.1	Skenario Uji coba <i>Recurisve Feature Elimination</i> . . . . .	31
4.3.2	Skenario uji coba <i>GridSearchCV</i> untuk <i>hyperparameter Support Vector Machine</i> . . . . .	31
4.3.3	Uji coba SVM dengan <i>Feature Selection</i> RFE dan tanpa RFE . . . . .	32
4.3.4	Uji coba dengan dataset lain . . . . .	35
BAB 5	SIMPULAN DAN SARAN . . . . .	36
5.1	Simpulan . . . . .	36
5.2	Saran . . . . .	36
	DAFTAR PUSTAKA . . . . .	37



## DAFTAR TABEL

Tabel 3.1	Dataset yang digunakan . . . . .	15
Tabel 4.1	Akurasi hyperparameter manual . . . . .	32
Tabel 4.2	Perbandingan SVM dengan RFE dan tanpa RFE . . . . .	34
Tabel 4.3	Perbandingan Summary Classification Matrix . . . . .	34
Tabel 4.4	Perbandingan feature selection dengan penelitian [?] . . . . .	34



## DAFTAR GAMBAR

Gambar 2.1	<i>Proses Recursive Feature Elimination</i>	6
Gambar 2.2	<i>Ilustrasi Support Vector Machine</i>	8
Gambar 2.3	<i>Ilustrasi Polynomial Kernel</i>	9
Gambar 2.4	<i>Ilustrasi Linear Kernel</i>	10
Gambar 2.5	<i>Ilustrasi RBF Kernel</i>	10
Gambar 2.6	<i>Classification Rate Table</i>	12
Gambar 3.1	<i>Flowchart Utama</i>	19
Gambar 3.2	<i>Flowchart PreProcessing</i>	21
Gambar 3.3	<i>Flowchart Recursive Feature Elimination</i>	22
Gambar 3.4	<i>Flowchart Support Vector Machine</i>	23
Gambar 4.1	<i>Peringkat fitur</i>	31
Gambar 4.2	<i>Hasil GridSearchCV</i>	31
Gambar 4.3	<i>Confussion Matrix data sebelum RFE</i>	33
Gambar 4.4	<i>Confussion Matrix data setelah RFE</i>	33
Gambar 4.5	<i>Hasil Evaluasi RFE</i>	35
Gambar 4.6	<i>Hasil GridSearchCV</i>	35

UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

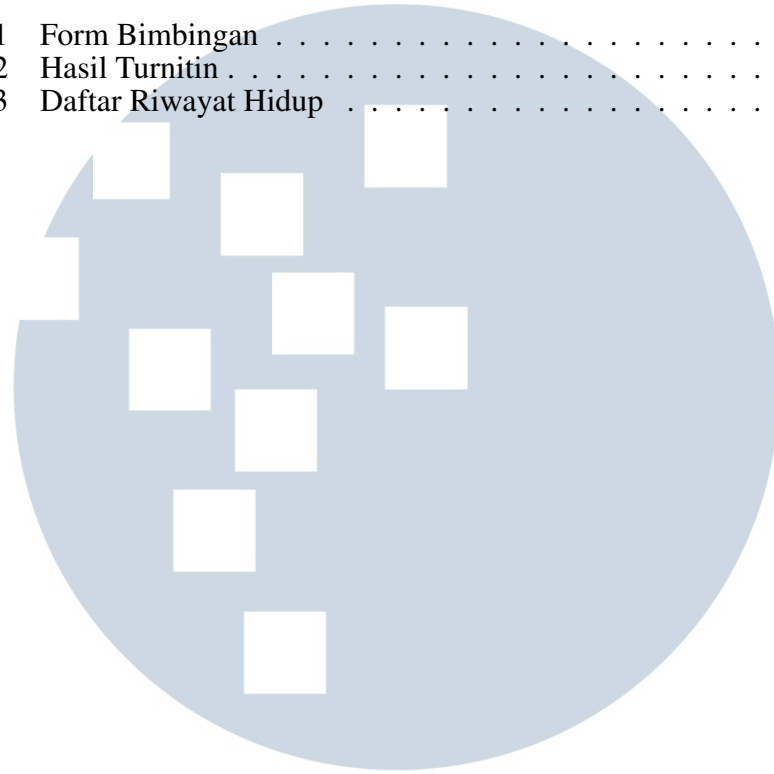
## Listings

4.1	Kode import library dan dataset . . . . .	26
4.2	Kode data cleaning . . . . .	26
4.3	Kode Standardization Data . . . . .	27
4.4	Kode label encoding . . . . .	28
4.5	Kode implementasi Recursive Feature Elimination model SVM . . . . .	28
4.6	Kode implementasi evaluasi model RFE . . . . .	29
4.7	Kode implementasi GridSearchCV . . . . .	29
4.8	Kode implementasi Support Vector Machine . . . . .	30
4.9	Kode implementasi Confussion Matrix . . . . .	30



## DAFTAR LAMPIRAN

Lampiran 1	Form Bimbingan . . . . .	38
Lampiran 2	Hasil Turnitin . . . . .	40
Lampiran 3	Daftar Riwayat Hidup . . . . .	47



# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA