

**IMPLEMENTASI *NATURAL LANGUAGE PROCESSING* DAN
ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK
MENDETEKSI URL *PHISHING* BERBASIS WEBSITE**



TUGAS AKHIR

**Disusun untuk Memenuhi Syarat Menyelesaikan Pendidikan
Sarjana Terapan pada Jurusan Teknik Elektro
Program Studi Teknik Telekomunikasi**

OLEH :

NABILA

062140352380

POLITEKNIK NEGERI SRIWIJAYA

PALEMBANG

2025

TUGAS AKHIR
IMPLEMENTASI *NATURAL LANGUAGE PROCESSING* DAN
ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK
MENDETEKSI URL *PHISHING* BERBASIS WEBSITE



Disusun Untuk Memenuhi Syarat Menyelesaikan Pendidikan
Sarjana Terapan Pada Jurusan Teknik Elektro
Program Studi Teknik Telekomunikasi

Oleh :

Nama : Nabila
Dosen Pembimbing 1 : Ir. Emilia Hesti, S.T., M.Kom.
Dosen Pembimbing II : Ir. Aryanti, S.T., M.Kom.

POLITEKNIK NEGERI SRIWIJAYA
PALEMBANG
2025

**IMPLEMENTASI *NATURAL LANGUAGE PROCESSING* DAN
ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK
MENDETEKSI URL *PHISHING* BERBASIS WEBSITE**



TUGAS AKHIR

Disusun untuk Memenuhi Syarat Menyelesaikan Pendidikan

Sarjana Terapan pada Jurusan Teknik Elektro

Program Studi Teknik Telekomunikasi

Politeknik Negeri Sriwijaya

OLEH:

NABILA

062140352380

Menyetujui,

Pembimbing I,

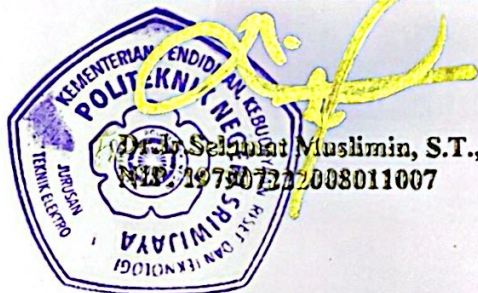
**Ir. Emilia Hesti, S.T., M.Kom.
NIP. 197205271998022001**

**Palembang, Agustus 2025
Pembimbing II,**

**Ir. Aryanti, S.T., M.Kom.
NIP. 197708092002122002**

Mengetahui,

Ketua Jurusan Teknik Elektro



**Dr. Ismail Muslimin, S.T., M.Kom., IPM.
NIP. 197507222008011007**

**Koordinator Program Studi
Sarjana Terapan Teknik Telekomunikasi**

**Mohammad Fadhli, S.Pd., M.T.
NIP. 199004032018031001**

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini menyatakan :

Nama : Nabila
Jenis Kelamin : Perempuan
Tempat, Tanggal Lahir : Palembang, 29 Maret
2004 Alamat : Lr Sepakat Rt 12 Rw 03
NPM : 062140352380
Program Studi : Sarjana Terapan Teknik Telekomunikasi
Jurusan : Teknik Elektro
Judul Skripsi / Laporan Akhir : Implementasi *Natural Language Processing* Dan Algoritma *Support Vector Machine* Untuk Mendeteksi Url *Phishing* Berbasis Website

Menyatakan dengan sesungguhnya bahwa:

1. Tugas Akhir ini adalah hasil karya saya sendiri serta bebas dari Tindakan plagiasi dan sumber – sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.
2. Dapat menyelesaikan segala urusan terkait pengumpulan revisi Tugas Akhir yang sudah disetujui oleh dewan penguji paling lama 1 bulan setelah ujian Tugas Akhir.
3. Dapat menyelesaikan segala urusan peminjaman/penggantian alat/buku dan lainnya paling lama 1 bulan setelah ujian Tugas Akhir.

Apabila dikemudian hari diketahui ada pernyataan yang terbukti tidak benar dan tidak dapat dipenuhi, maka saya siap bertanggung jawab dan menerima sanksi tidak diikutsertakan dalam prosesi wisuda serta dimasukkan dalam daftar hitam oleh jurusan Teknik Elektro sehingga berdampak tertundanya pengambilan Ijazah & Transkrip (ASLI & COPY). Demikian surat pernyataan ini dibuat dengan sebenar- benarnya dan dalam keadaan sadar tanpa paksaan.

Palembang, Agustus 2025

Yang Menyatakan



(Nabila)

MOTTO DAN PERSEMBAHAN

“Libatkanlah Allah dalam Segala Urusan, Agar yang Berat Menjadi Ringan dan yang Sulit Menjadi Mudah”

“Hai orang-orang yang beriman, mintalah pertolongan kepada Allah dengan sabar dan salat. Sesungguhnya Allah beserta orang-orang yang sabar”

-- Q.S Al-Baqarah:15 –

“Selama masih ada yang Namanya perjuangan, tidak ada yang sia-sia”

-- R.A. Kartini –

Tugas Akhir ini kupersembahkan kepada :

1. Allah SWT
2. Ayah, Mama, Kakak, dan seluruh keluarga besar yang selalu memberikan do'a, cinta, kasih sayang, serta dukungan yang tak pernah putus.
3. Ibu Ir.Emilia Hesti, S.T., M.Kom dan Ibu Ir.Aryanti, S.T., M.Kom selaku dosen pembimbing. Terima kasih yang sebesar-besarnya atas segala bimbingan, arahan, serta dukungan yang telah diberikan selama proses penyusunan tugas akhir ini.
4. Teman-teman seperjuangan TEB 2021. Terima kasih atas kebersamaan, semangat, dan dukungannya selama masa perkuliahan. Semoga kita semua terus melangkah maju dan meraih cita-cita bersama.

ABSTRAK

IMPLEMENTASI *NATURAL LANGUAGE PROCESSING* DAN ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK MENDETEKSI URL *PHISHING*

(2025:xv + 60 halaman + 36 gambar +17 tabel + lampiran)

Nabila

062140352380

JURUSAN TEKNIK ELEKTRO

**PROGRAM STUDI SARJANA TERAPAN TEKNIK TELEKOMUNIKASI
POLITEKNIK NEGERI SRIWIJAYA**

Phishing merupakan bahaya yang signifikan dalam keamanan siber, menggunakan URL jahat untuk menyesatkan pengguna agar mengungkapkan informasi penting. Penelitian ini berupaya untuk membuat model deteksi URL *phishing* menggunakan pembelajaran mesin melalui integrasi ekstraksi fitur URL struktural, metodologi *Natural Language Processing* (NLP), dan algoritma klasifikasi *Support Vector Machine* (SVM). Indikator tren *phishing* diperoleh dari fitur-fitur seperti panjang URL, jumlah titik, dan garis miring, sementara konten URL dikuantifikasi sebagai vektor numerik menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Semua karakteristik selanjutnya diintegrasikan sebagai input ke dalam model *support vector machine* dengan kernel linier untuk klasifikasi. Hasil evaluasi dari laporan klasifikasi menunjukkan bahwa integrasi TF-IDF dan SVM kernel linier mencapai kinerja optimal, dengan akurasi 90%, presisi 92%, *recall* 89%, dan F1-score 90%. Sebaliknya, *convusion matrix* menunjukkan akurasi 90,29%, presisi 91,66%, ingatan 88,62%, dan F1-Score 90,12%. Studi ini terutama berkontribusi dengan mengintegrasikan NLP dan SVM ke dalam model deteksi phishing adaptif terpadu melalui penggabungan aspek struktural dan tekstual URL. Strategi ini memfasilitasi deteksi *phishing* yang lebih baik dibandingkan dengan teknik yang hanya bergantung pada karakteristik manual. Model ini, tidak seperti penelitian lain yang difokuskan pada contoh tertentu atau NLP yang dikecualikan, dirancang untuk mengidentifikasi banyak kategori URL *phishing* secara luas, sehingga meningkatkan relevansinya dalam menangani serangan siber yang terus berkembang.

Keyword: *Phishing; Natural Language Processing; Support Vector Machine; Term Frequency-Inverse Document Frequency.*

ABSTRACT

IMPLEMENTATION OF NATURAL LANGUAGE PROCESSING AND SUPPORT VECTOR MACHINE ALGORITHM FOR PHISHING URL DETECTION

(2025:xv + 60 pages + 36 pictures + 17 tabel + appendices)

Nabila

062140352380

ELECTRICAL ENGINEERING DEPARTMENT

PROGRAM OF STUDY IN APPLIED GRADUATION OF THE

TELECOMMUNICATION ENGINEERING

STATE POLYTECHNIC OF SRIWIJAYA

Phishing is a significant threat in cybersecurity, using malicious URLs to deceive users into disclosing sensitive information. This study aims to develop a phishing URL detection model using machine learning through the integration of structural URL feature extraction, Natural Language Processing (NLP) methods, and the Support Vector Machine (SVM) classification algorithm. Phishing trend indicators are obtained from features such as URL length, number of dots, and slashes, while the URL content is quantified as numerical vectors using Term Frequency-Inverse Document Frequency (TF-IDF). All characteristics are integrated as input into a linear kernel support vector machine model for classification. Evaluation results from the classification report indicate that the integration of TF-IDF and linear kernel SVM achieves optimal performance, with 90% accuracy, 92% precision, 89% recall, and a 90% F1-score. Conversely, the confusion matrix shows 90.29% accuracy, 91.66% precision, 88.62% recall, and a 90.12% F1-score. This study contributes significantly by integrating NLP and SVM into a unified adaptive phishing detection model through a combination of URL structural and textual aspects. This strategy enables better phishing detection compared to techniques that rely solely on manual features. Unlike previous studies that focus on specific cases or exclude NLP, this model is designed to identify a wide range of phishing URL categories, enhancing its relevance in addressing evolving cyber threats.

***Keywords : Phishing; Natural Language Processing; Support Vector Machine;
Term Frequency-Inverse Document Frequency***

KATA PENGANTAR

Segala puji dan syukur dipanjatkan ke hadirat Allah SWT, karena atas rahmat dan hidayah-Nya, penulis dapat menyelesaikan laporan tugas akhir yang berjudul **“IMPLEMENTASI *NATURAL LANGUAGE PROCESSING* DAN ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK MENDETEKSI URL *PHISING* BERBASIS WEBSITE”**. Tugas akhir ini disusun sebagai salah satu syarat untuk menyelesaikan studi di Jurusan Teknik Elektro, Program Studi Sarjana Terapan Teknik Telekomunikasi, Politeknik Negeri Sriwijaya, Palembang.

Dalam proses penyusunan laporan ini, penulis mengucapkan terima kasih kepada Ibu Ir.Emilia Hesti, S.T., M.Kom. dan Ibu Ir.Aryanti, S.T., M.Kom. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta masukan yang sangat berarti dalam penyelesaian laporan tugas akhir ini. Selain itu, penulis juga ingin menyampaikan rasa terima kasih kepada:

- 1.Orang Tua dan kakak penulis yang selalu memberikan dukungan, doa dan bantuan kepada penulis selama melaksanakan kegiatan Tugas Akhir.
2. Bapak Ir. Irawan Rusnadi, M.T. selaku Direktur Politeknik Negeri Sriwijaya.
3. Bapak Dr.Ir.Selamat Muslimin, S.T., M.Kom., IPM. selaku Ketua Jurusan Teknik Elektro Politeknik Negeri Sriwijaya.
4. Ibu Lindawati, ST,M.T.I selaku Sekretaris Jurusan Teknik Elektro Politeknik Negeri Sriwijaya.
5. Bapak Mohammad Fadhli,S.Pd.,M.T selaku Koordinator Program Studi Sarjana Terapan Teknik Telekomunikasi Politeknik Negeri Sriwijaya.
6. Bapak/Ibu dosen, karyawan, dan seluruh staff Program Studi Teknik Telekomunikasi.
7. Teman Kelas TEB khususnya teman terdekat penulis yaitu Siska, Luthfia, Sulis dan Kiara yang sudah memberikan penulis motivasi dan dukungan dalam menyelesaikan laporan Tugas Akhir.
8. Rama Alfajri yang telah memberikan bantuan, support dan motivasi untuk penulis dalam menyelesaikan laporan Tugas Akhir.

9. Kak Rahmat, kak johan, kak puspa yang sudah memberikan bantuan dan support untuk penulis dalam menyelesaikan laporan tugas akhir.
10. Seluruh pihak yang telah memberikan bantuan yang tidak bisa penulis sebutkan satu persatu, baik secara langsung maupun tidak langsung.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi penyusunan laporan yang lebih baik di masa depan. Akhir kata, semoga laporan ini dapat bermanfaat bagi semua pihak yang membacanya.

Palembang, Agustus 2025

Penulis

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI.....	iii
DAFTAR GAMBAR	vi
DAFTAR TABEL.....	viii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	4
1.4 Tujuan.....	4
1.5 Manfaat.....	5
1.6 Metode Penelitian.....	5
1.7 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 <i>Phishing</i>	8
2.2 <i>Uniform Resource Locator (URL)</i>	9
2.3 <i>Artificial Intelligence (AI)</i>	13
2.4 <i>Machine Learning (ML)</i>	13
2.5 <i>Natural Language Processing (NLP)</i>	15
2.6 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	16
2.7 <i>Support Vector Machine (SVM)</i>	18
2.8 Evaluasi	21
2.9 Website	22
2.10 Visual Studio Code.....	24
2.11 Flask.....	24

2.12 Google Colab.....	25
2.13 Perbandingan Penelitian	26
BAB III METODOLOGI PENELITIAN.....	30
3.1 Kerangka Penelitian.....	30
3.2 Pengumpulan Dataset dan Pemrosesan Dataset	31
3.3 Pelatihan Algoritma <i>Support Vector Machine</i> pada Dataset.....	32
3.4 Perancangan Sistem.....	34
3.5 <i>Flowchart</i> Sistem.....	35
3.6 Blok Diagram	37
3.7 Perancangan Perangkat Lunak.....	38
3.8 Integrasi Antara Website dengan Algoritma <i>Support Vector Machine</i> (SVM) dan <i>Natural Language Processing</i> (NLP)	39
BAB IV HASIL DAN PEMBAHASAN	40
4.1 <i>Preprocessing</i> Url.....	40
4.2 Ekstraksi Fitur Url	42
4.3 Ekstraksi Fitur.....	44
4.4 Evaluasi Model	48
4.4.1 <i>Classification Report</i>	48
4.4.2 <i>Confusion Matrix</i>	49
4.5 Tampilan Website.....	51
4.6 Uji coba.....	53
4.7 Hasil Perbandingan.....	55
BAB V KESIMPULAN DAN SARAN.....	59
5.1 Kesimpulan.....	59
5.2 Saran	59

DAFTAR PUSTAKA	61
----------------------	----

DAFTAR GAMBAR

Gambar 2. 1 <i>Phishing</i>	9
Gambar 2. 2 Protokol	11
Gambar 2. 3 <i>Domain Name System</i>	11
Gambar 2. 4 <i>Path</i>	11
Gambar 2. 5 <i>Query</i>	12
Gambar 2. 6 UTM.....	12
Gambar 2. 7 Anchor.....	12
Gambar 2. 8 Elemen URL.....	12
Gambar 2. 9 <i>Artificial Intelligence</i>	13
Gambar 2. 10 <i>Machine Learning</i>	15
Gambar 2. 11 <i>Natural Language Processing</i>	16
Gambar 2. 12 Cara kerja TF-IDF	17
Gambar 2. 13 <i>Support vector machine</i>	18
Gambar 2. 14 Cara kerja svm.....	19
Gambar 2. 15 Cara Kerja SVM.....	20
Gambar 2. 16 Linear separable dan non linear separable	21
Gambar 2. 17 Website.....	23
Gambar 2. 18 Visual Studio Code.....	24
Gambar 2. 19 Logo Flask.....	25
Gambar 2. 20 Logo Google Colab	26
Gambar 3. 1 Kerangka penelitian.....	30
Gambar 3. 2 <i>Hyperlane SVM</i>	32
Gambar 3. 3 Perancangan sistem	34
Gambar 3. 4 <i>flowchart</i>	36
Gambar 3. 5 Blok diagram.....	37
Gambar 3. 6 Tampilan web pendeteksi url <i>phishing</i>	38
Gambar 4. 1 Distribusi Data.....	40
Gambar 4. 2 Source code preprocessing URL	41
Gambar 4. 3 Source code fitur lainnya.....	42

Gambar 4. 4 <i>Source code</i> ekstraksi fitur menggunakan <i>Tfidf Vectorizer</i>	45
Gambar 4. 5 Hasil <i>classification report</i>	48
Gambar 4. 6 Hasil <i>confusion matrix</i>	49
Gambar 4. 7 Tampilan <i>home</i> pada website	51
Gambar 4. 8 Tampilan <i>about</i> pada website.....	51
Gambar 4. 9 Uji coba url pada website	52
Gambar 4. 10 Hasil yang ditampilkan pada website.....	52

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	26
Tabel 4. 1 Sample data	40
Tabel 4. 2 Hasil <i>Preprocessing</i> Data.....	41
Tabel 4. 3 Hasil Fitur Url	43
Tabel 4. 4 Hasil token dari kedua sample url	46
Tabel 4. 5 Perhitungan <i>Term Frequency</i>	46
Tabel 4. 6 Hasil $df(t)$	46
Tabel 4. 7 Hasil Perhitungan IDF	47
Tabel 4. 8 Perhitungan TF-IDF	47
Tabel 4. 9 Hasil <i>Term Frequency-Inverse Document Frequency</i>	47
Tabel 4. 10 Komponen penting dari <i>confusion matrix</i>	49
Tabel 4. 11 <i>Blackbox Testing</i>	53
Tabel 4. 12 Hasil pengujian data baru.....	54
Tabel 4. 13 Hasil perbandingan <i>classification report</i> dari kombinasi SVM dengan dan tanpa NLP.....	56
Tabel 4. 14 Hasil perbandingan <i>confusion matrix</i> kombinasi SVM dengan dan tanpa NLP	56
Tabel 4. 15 <i>Confusion matrix</i> svm tanpa nlp	56
Tabel 4. 16 Grafik perbandingan <i>classification report</i> SVM dengan dan tanpa NLP	57