

**DETEKSI *WEBSITE PHISING* MENGGUNAKAN TEKNIK
DATA MINING DAN WEB CRAWLING**

TUGAS AKHIR

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana
Program Studi S-1 Teknologi Informasi



Disusun oleh:

NOVIANA WULANDARI
212104009

**PROGRAM STUDI S-1 TEKNOLOGI INFORMASI
FAKULTAS TEKNIK & TEKNOLOGI INFORMASI
UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA
2025**

HALAMAN PENGESAHAN

TUGAS AKHIR
DETEKSI WEBSITE PHISING MENGGUNAKAN TEKNIK DATA
MINING DAN WEB CRAWLING

Diajukan oleh:

NOVIANA WULANDARI

212104009

Telah dipertahankan di depan dewan penguji dan dinyatakan sah
sebagai salah satu syarat untuk memperoleh gelar Sarjana
di Fakultas Teknik & Teknologi Informasi
Universitas Jenderal Achmad Yani Yogyakarta

Tanggal: 18 Juli 2025

Mengesahkan:

Pembimbing

Penguji


Chanief Budi Setiawan, S.T., M.Eng.

NIDN: 0514068101


Alfirna Rizqi Lahitani, S.Kom., M.Eng.

NIDN: 0506019202

Ketua Program Studi S-1 Teknologi Informasi
Fakultas Teknik & Teknologi Informasi
Universitas Jenderal Achmad Yani Yogyakarta




Rama Sahtyawan, S.T., M.Cs.

NPP: 2019.13.00150

PERNYATAAN

Saya yang bertanda tangan di bawah ini, adalah mahasiswa Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta,

Nama : Noviana Wulandari
NPM : 212104009
Program Studi : S-1 Teknologi Informasi
Judul Tugas Akhir : Deteksi *Website Phising* Menggunakan Teknik Data Mining dan *Web Crawling*

Menyatakan bahwa hasil penelitian dengan judul tersebut di atas adalah asli karya saya sendiri dan bukan hasil plagiarisme. Semua referensi dan sumber terkait yang dikutip dalam karya ilmiah ini telah ditulis sesuai kaidah penulisan ilmiah yang berlaku. Dengan ini, saya menyatakan untuk menyerahkan hak cipta penelitian kepada Universitas Jenderal Achmad Yani Yogyakarta guna kepentingan ilmiah.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya tanpa ada paksaan dari pihak mana pun. Apabila terdapat kekeliruan atau ditemukan adanya pelanggaran akademik di kemudian hari, maka saya bersedia menerima konsekuensi yang berlaku sesuai ketentuan akademik.

Yogyakarta, 18 Juli 2025



Noviana Wulandari

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul: “Deteksi *Website Phising* Menggunakan Teknik *Data Mining* dan *Web Crawling*”. Penyusunan laporan ini merupakan salah satu persyaratan untuk menyelesaikan studi di Program Studi S-1 Teknologi Informasi Fakultas Teknik & Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta. Laporan ini dapat diselesaikan atas bimbingan, arahan, dan bantuan dari berbagai pihak. Pada kesempatan ini penulis dengan rendah hati mengucapkan terima kasih dengan setulus-tulusnya kepada:

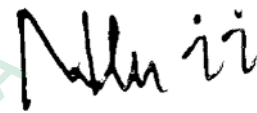
1. Kepada kedua orang tua tercinta (Bapak Safrudin & Ibu Riyanah) orang hebat yang selalu menjadi penyemangat penulis serta sandaran dari kerasnya kehidupan, terimakasih sudah berjuang untuk kehidupan penulis, beliau memang tidak sempat merasakan pendidikan sampai bangku perkuliahan, namun beliau mampu mendidik penulis, memotivasi, memberikan dukungan hingga penulis mampu menyelesaikan studinya sampai sarjana dengan tepat waktu;
2. Kepada saudara penulis Safira Ayu Lestari, S.Pd. Terimakasih atas dukungan secara moril maupun materi, terimakasih juga atas segala motivasi dan dukungan yang diberikan kepada penulis sehingga penulis mampu menyelesaikan studi tepat waktu sampai sarjana;
3. Bapak Rama Sahtywan, S.T., M.Cs. selaku Ketua Program Studi S-1 Teknologi Informasi Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
4. Bapak Chanief Budi Setiawan, S.T., M.Eng. selaku Dosen Pembimbing Tugas Akhir yang telah membantu selama berjalannya penelitian hingga penyusunan laporan;

5. Ibu Alfirma Rizki Lahitani, S.Kom., M.Eng. selaku Dosen Pembimbing Akademik yang telah membimbing selama masa perkuliahan berlangsung hingga saat ini;
6. Para dosen yang telah memberikan banyak bekal ilmu pengetahuan kepada penulis selama menjadi mahasiswa di Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
7. Kepada keponakan-keponakan tersayang Rahel Adreena Saila dan Radha Aiko Mahalini Elaksi, terimakasih atas kelucuan kalian yang membuat penulis semangat dan selalu membuat penulis senang, sehingga penulis penulis semangat untuk mengerjakan tugas akhir ini sampai selesai;
8. Teruntuk Bella Dwi Utami, sahabat tersayang penulis yang selalu menemani, membantu, memberi motivasi dan semangat yang luar biasa. Terimakasih sudah menjadi sahabat yang sangat baik bahkan seperti saudara sendiri karena tidak pernah meninggalkan penulis sendirian, selalu menjadi garda terdepan saat penulis membutuhkan bantuan serta mendengarkan keluh kesah penulis selama diperantauan ini;
9. Teruntuk pemilik NIP 200006192021031183BLUD penulis ucapkan terimakasih telah menjadi bagian dari perjalanan hidup penulis, selalu ada dan tak henti-hentinya memberikan semangat dan dukungan serta mendengarkan keluh kesah penulis selama penyusunan tugas akhir ini;
10. Keluarga dan sahabat-sahabat penulis yang senantiasa memberikan doa, dukungan dan semangat disetiap perjuangan sehingga penulis dapat menyelesaikan tugas akhir ini dengan tepat waktu;
11. Teman-teman seperjuangan mahasiswa Prodi S-1 Teknologi Informasi di Universitas Jenderal Achmad Yani Yogyakarta yang sudah memberi dukungan dan kerja sama selama pembuatan tugas akhir;
12. *Lust but not least!*, terimakasih kepada wanita sederhana yang memiliki impian besar, namun terkadang sulit dimengerti isi kepalanya, yaitu penulis sendiri Noviana Wulandari. Seorang anak bungsu yang berjalan memasuki usia 23 tahun, sangat keras kepala dan yang penuh ambisi, namun sifatnya seperti anak kecil. Terimakasih telah berusaha keras untuk meyakinkan dan

menguatkan diri sendiri bahwa kamu dapat menyelesaikan studi ini sampai selesai. Berbahagialah selalu rayakan kehadiranmu sebagai berkah dimana pun kamu menjejakkan kaki. Semoga langkah kebaikan selalu menyertaimu, dan semoga Allah selalu meridhoi setiap langkahmu serta menjagamu dalam lindungan-Nya.

Penulis menyadari bahwa laporan tugas akhir ini masih jauh dari kata sempurna. Maka dari itu dengan segala kerendahan hati penulis sangat menghargai adanya kritik dan saran yang membangun dari semua pihak yang bersedia meluangkan waktu untuk membaca laporan tugas akhir ini.

Yogyakarta, 18 Juli 2025



Noviana Wulandari

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
PERNYATAAN.....	iii
KATA PENGANTAR.....	iv
DAFTAR ISI.....	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN	xi
DAFTAR SINGKATAN.....	xii
INTISARI	xiv
ABSTRACT	xv
Bab 1 Pendahuluan	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah	3
1.3 Pertanyaan Penelitian.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Hasil Penelitian	4
Bab 2 Tinjauan Pustaka dan Landasan Teori.....	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori.....	8
2.2.1 <i>Phising</i>	8
2.2.2 <i>Data Mining</i>	9
2.2.3 <i>Web Crawling</i>	9
2.2.4 <i>PhisTank</i>	10
2.2.5 <i>Decision Tree</i>	10
2.2.6 <i>Classification and Regression Tree (CART)</i>	10
2.2.7 <i>Confusion Matrix</i>	11
Bab 3 Metode Penelitian	13

3.1	Bahan dan Alat Penelitian.....	16
3.2	Jalan Penelitian.....	17
Bab 4	Hasil Penelitian	19
4.1	Ringkasan Hasil.....	19
4.2	Pengumpulan data	20
4.3	<i>Pre-Processing</i> Data	21
4.4	Ekstraksi Fitur	22
4.5	Normalisasi Data	24
4.6	Pelatihan dan Pengujian Model <i>Decision Tree CART</i>	25
4.7	Evaluasi Model <i>Decision Tree CART</i>	26
4.7.1	ROC Curve & AUC	29
4.7.2	<i>Precision-Recall</i> Curve	30
4.8	Cross validation dan Evaluasi Fairness	31
4.9	Visualisasi Pohon Keputusan.....	32
4.10	Uji Prediksi <i>Real-Time</i>	35
Bab 5	Kesimpulan dan Saran.....	38
5.1.	Kesimpulan	38
5.2.	Saran	39
DAFTAR PUSTAKA		40
LAMPIRAN.....		42

DAFTAR TABEL

Tabel 2.1 Ringkasan Penelitian Terdahulu.....	7
Tabel 2.2 Komponen Klasifikasi Biner	11
Tabel 4.1 Daftar URL yang Digunakan	22
Tabel 4.2 <i>Confusion Matrix</i>	26
Tabel 4.3 <i>Classification Report</i>	28
Tabel 4.4 Hasil Cross-Validation 5-Fold	32
Tabel 4.5 Akurasi Model pada Data Latih dan Uji	32
Tabel 4.6 Prediksi <i>Real-Time</i>	36

DAFTAR GAMBAR

Gambar 2.1 Jenis Serangan Phising	8
Gambar 3.1 Alur Tahapan Sistematis.....	13
Gambar 4.1 Sampel Dataset <i>Phising</i>	21
Gambar 4.2 Sampel Dataset <i>Legitimate</i>	21
Gambar 4.3 Normalisasi dengan MinMax Scaler	24
Gambar 4.4 Data Latih	25
Gambar 4.5 Data Uji.....	25
Gambar 4.6 Visualisasi Confusion Matrix CART	27
Gambar 4.7 Kurva ROC AUC.....	30
Gambar 4.8 Kurva <i>Precision-Recall</i>	31
Gambar 4.9 <i>Decision Tree</i>	34
Gambar 4.10 <i>Pie Chart</i> Distribusi Prediksi <i>Real-Time</i>	37

DAFTAR LAMPIRAN

Lampiran 1 <i>Script</i> Kode	43
Lampiran 2 Hasil Ekstraksi Fitur	45
Lampiran 3 Jadwal Penelitian	46
Lampiran 4 Kartu Bimbingan Tugas Akhir	47
Lampiran 5 Hasil Cek Plagiarisme.....	48

PERPUSTAKAAN
UNIVERSITAS JENDRAL ACHMAD YANI
YOGYAKARTA

DAFTAR SINGKATAN

AUC	<i>Area Under Curve</i>
API	<i>Application Programming Interface</i>
APWG	<i>Anti-Phising Working Group</i>
BBSN	<i>Badan Siber dan Sandi Negara</i>
CART	<i>Classification and Regression Tree</i>
CSV	<i>Comma Separated Values</i>
FBI	<i>Federal Bureau of Investigation</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
FPR	<i>False Positive Rate</i>
F1-Score	<i>F-Measure (Harmonic Mean of Precision & Recall)</i>
GCP	<i>Google Colaboratory (Google Colab)</i>
GUI	<i>Graphical User Interface</i>
HTML	<i>Hypertext Markup Language</i>
HTTPS	<i>Hypertext Transfer Protocol Secure</i>
IDS	<i>Intrusion Detection System</i>
IP	<i>Internet Protocol</i>
MAE	<i>Mean Absolute Error</i>
MSE	<i>Mean Squared Error</i>
ML	<i>Machine Learning</i>
PR Curve	<i>Precision-Recall Curve</i>
ROC	<i>Receiver Operating Characteristic</i>
SVM	<i>Support Vector Machine</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>
TPR	<i>True Positive Rate</i>
UAWY	<i>Universitas Adiwira Yogyakarta</i>
UCI	<i>University of California, Irvine</i>

URL	<i>Uniform Resource Locator</i>
WHOIS	<i>World Host Information System</i>
XXS	<i>Cross Site Scripting</i>

PERPUSTAKAAN
UNIVERSITAS JENDRAL ACHMAD YANI
YOGYAKARTA