ANALISIS KLASIFIKASI URL *PHISHING* MENGGUNAKAN *RANDOM FOREST* BERDASARKAN FITUR URL

SKRIPSI

Untuk Memenuhi Persyaratan Mencapai Gelar Sarjana Komputer (S.Kom)
Pada Program Studi Teknologi Informasi Fakultas Teknik UM Palembang

Disusun Oleh :

Syahrul Arya Ramadhan
NIM : 162022065

PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH PALEMBANG

2026

ANALISIS KLASIFIKASI URL *PHISHING* MENGGUNAKAN *RANDOM FOREST* BERDASARKAN FITUR URL

SKRIPSI

Untuk Memenuhi Persyaratan Mencapai Gelar Sarjana Komputer (S.Kom)
Pada Program Studi Teknologi Informasi Fakultas Teknik UM Palembang

Disusun Oleh :

Syahrul Arya Ramadhan
NIM : 162022065

PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH PALEMBANG

2026

HALAMAN PENGESAHAN

ANALISIS KLASIFIKASI URL *PHISHING* MENGGUNAKAN *RANDOM FOREST* BERDASARKAN FITUR URL

Oleh:

Syahrul Arya Ramadhan

162022065

Menyetujui,

Dosen pembimbing utama	**Dosen pendamping**
Kms. M Wahyu Hidayat, S.Kom., M.Kom	**Dr. Lucky Indra kesuma, S.SI., M.Kom**
NBM/NIDN : 1255881/0225068904	**NBM/NIDN : 1582348/0225099002**

Disetujui, **Dekan Fakultas Tekník**	**Program Studi Teknologi informasi** **Ketua Program Studi**
Ir. A. Junaidi, M.T	**Karnadi, S.Kom., M.kom**
NBM/NIDN : 763050/0202026502	**NBM/NIDN : 1088893/0210038202**

HALAMAN PERSETUJUAN

Judul Skripsi Penelitian : Analisis Klasifikasi URL *Phishing* menggunakan *Random Forest* berdasarkan fitur URL

Oleh **Syahrul Arya Ramadhan** NIM **162022065** Skripsi ini telah disetujui dan disahkan oleh Tim Penguji Program Studi Teknologi Informasi Konsentrasi Manajemen Tata Kelola Program Strata 1 Universitas Muhammadiyah Palembang pada 27 April 2026 dan telah Dinyatakan LULUS

Mengetahui,

Ketua Program Studi Teknologi Informasi
Universitas Muhammadiyah Palembang

Karnadi, S. Kom., M. Kom
NBM/NIDN: 1088893/0210038202

Tim Penguji

Ketua Penguji

Kms. M Wahyu Hidayat, S.Kom., M.Kom
NBM/NIDN: 1255881/0225068904

Penguji 1,

Karnadi, S. Kom., M. Kom
NBM/NIDN: 1088893/0210038202

Penguji 2,

Jimmie, S.Kom., M.Kom
NBM/NIDN: 1340253/0222047702

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama	: Syahrul Arya Ramadhan
NIM	: 162022065
Fakultas/Prodi	: Teknik / Teknologi Informasi
Judul Skripsi	: Analisis Klasifikasi URL *Phishing* menggunakan *Random Forest* berdasarkan fitur URL

Dengan ini saya menyatakan dengan sebenar-benarnya bahwa skripsi ini merupakan hasil karya asli saya sendiri dan bukan merupakan praktik plagiarisme, pencurian hasil karya milik orang lain, maupun hasil kerja pihak lain yang dilakukan demi kepentingan saya, baik karena hubungan material maupun non-material. Seluruh isi dalam skripsi ini merupakan karya tulis yang bersifat orisinal dan autentik.

Apabila di kemudian hari ditemukan bukti yang kuat mengenai adanya ketidaksesuaian antara pernyataan ini dengan fakta yang ada, saya bersedia menanggung segala konsekuensi hukum dan diproses oleh tim verifikasi fakultas yang berwenang, dengan sanksi terberat berupa pembatalan status kelulusan serta gelar kesarjanaan saya.

Demikian surat pernyataan ini saya buat dengan penuh kesadaran dan tanpa ada tekanan maupun paksaan dari pihak manapun, semata-mata demi menegakkan integritas akademik di institusi ini.

Hormat Saya,

Syahrul Arya Ramadhan

MOTTO DAN PERSEMBAHAN

MOTTO

"Tiada kekayaan yang lebih utama dari pada akal, tidak ada keadaan yang lebih menyedihkan daripada kebodohan, dan tiada warisan yang lebih indah daripada pendidikan."

- Ali Bin Abi Thalib

"Disiplin tidak butuh motivasi, disiplin artinya bergerak dengan keinginan sendiri, disiplin artinya tetap bergerak sekalipun lelah."

- Khabib Nurmagomedov

PERSEMBAHAN

Dengan penuh rasa syukur atas segala nikmat dan kekuatan yang diberikan oleh Allah SWT, serta dengan kerendahan hati, penulis mempersembahkan karya sederhana ini kepada:

1. Kepada kedua orang tua tercinta, sumber inspirasi dan alasan utama bagi penulis untuk tetap tegak berdiri menghadapi setiap rintangan. Terima kasih sedalam-dalamnya atas kasih sayang yang tak bertepi, doa yang tak putus dalam sujud, serta pengorbanan tanpa batas yang senantiasa menjadi cahaya di setiap fase perjuangan penulis. Karya ini adalah persembahan kecil atas asa dan amanah yang kalian titipkan. Semoga pencapaian ini menjadi awal bagi penulis untuk terus membahagiakan dan membalas ketulusan yang telah kalian berikan.
2. Untuk kakak tersayang, yang senantiasa memberikan doa dan semangat dalam setiap langkah perjalanan penulis.

3. Kepada Bapak Kemas Muhammad Wahyu Hidayat, S.Kom., M.Kom. selaku Dosen Pembimbing Utama, serta Bapak Dr. Lucky Indra Kesuma, S.Si., M.Kom. selaku Dosen Pembimbing Pendamping. Terima kasih yang sebesar-besarnya atas segala dedikasi, bimbingan, serta wawasan intelektual yang telah diberikan. Arahan dan motivasi yang Bapak berikan tidak hanya membantu penulis dalam menyelesaikan skripsi ini dengan baik, tetapi juga telah memperluas cakrawala berpikir penulis dalam memahami bidang ilmu ini secara lebih mendalam.

4. Teman-teman seperjuangan angkatan 2022 Teknologi Informasi, yang telah membersamai dalam suka, duka, dan berbagai cerita selama di bangku perkuliahan.

5. Almamater tercinta Universitas Muhammadiyah Palembang, yang telah menjadi tempat penulis tumbuh, belajar, dan meraih pengalaman berharga hingga mencapai tahap penyelesaian studi ini.

6. Terakhir Untuk diri sendiri, Syahrul Arya Ramadhan, terima kasih telah bertahan sejauh ini. Terima kasih karena tetap kuat di tengah lelah, tidak menyerah saat keadaan terasa sulit, dan terus melangkah melewati setiap proses yang tidak mudah. Segala perjuangan, kesabaran, dan usaha yang telah dilalui akhirnya membuahkan hasil hingga skripsi ini dapat terselesaikan. Semoga ini menjadi awal dari langkah-langkah besar berikutnya.

ABSTRAK

Serangan Serangan *Phishing* merupakan salah satu bentuk kejahatan siber yang paling sering terjadi dan umumnya memanfaatkan tautan atau URL palsu untuk menipu korban agar memasukkan informasi sensitif. Modus *Phishing* terus berkembang mengikuti tren layanan digital, sehingga banyak URL *Phishing* dibuat menyerupai domain resmi dan sulit dikenali secara manual. Penelitian sebelumnya menekankan bahwa peningkatan *security awareness* merupakan langkah penting untuk mengurangi risiko *Phishing*. Namun, pendekatan kesadaran pengguna saja sering belum cukup karena variasi URL *Phishing* semakin kompleks dan dapat mengecoh pengguna awam maupun pengguna yang sudah cukup berpengalaman [1]. Oleh karena itu, diperlukan pendekatan teknis berbasis komputasi yang dapat membantu mengidentifikasi URL *Phishing* secara otomatis. Penelitian ini bertujuan untuk menganalisis klasifikasi URL *Phishing* menggunakan *Random Forest* berdasarkan fitur URL. Dataset yang digunakan berasal dari Kaggle dengan nama *Phishing URL Website*, yang berisi atribut-atribut seperti domain, *Top Level Domain* (TLD), jumlah karakter khusus pada URL (*NoOfOtherSpecialCharsInURL*), penggunaan protokol keamanan (*HTTPS*), jumlah baris kode halaman (*LineOfCode*), judul halaman (*Title*), serta tingkat kesesuaian antara domain dan judul (*DomainTitleMatchScore*) dan antara URL dan judul (*URLTitleMatchScore*), serta label *Phishing* (0/1). Proses penelitian dilakukan melalui tahapan *preprocessing* data, pembagian data latih dan data uji, pelatihan model *Random Forest*, serta evaluasi performa menggunakan metrik *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score* dengan bantuan Python. Hasil penelitian diharapkan mampu memberikan gambaran performa *Random Forest* dalam membedakan URL *Phishing* dan URL *legitimate* berdasarkan kombinasi fitur URL dan karakteristik halaman web. Temuan ini juga dapat menjadi referensi pengembangan sistem deteksi *Phishing* sebagai bentuk dukungan pencegahan *Phishing* dari sisi teknologi.

Kata kunci: *Phishing*, URL *Phishing, Random Forest, Machine learning*, Python.

ABSTRACT

Phishing attacks are one of the most common forms of cybercrime, typically utilizing fake links or URLs to trick victims into entering sensitive information. Phishing methods continue to evolve following digital service trends, resulting in many Phishing URLs being created to resemble official domains and difficult to identify manually. Previous research has emphasized that increasing security awareness is a crucial step in reducing the risk of Phishing. However, user awareness alone is often insufficient because Phishing URL variations are increasingly complex and can deceive both novice and experienced users [1]. Therefore, a computational-based technical approach is needed that can help identify Phishing URLs automatically. This study aims to analyze the classification of Phishing URLs using Random Forest based on URL features. The dataset used comes from Kaggle and is called Phishing URL Website. It contains attributes such as domain, Top Level Domain (TLD), number of special characters in the URL (NoOfOtherSpecialCharsInURL), security protocol usage (HTTPS), number of lines of page code (LineOfCode), page title (Title), the level of match between domain and title (DomainTitleMatchScore) and between URL and title (URLTitleMatchScore), and the Phishing label (0/1). The research process was carried out through data preprocessing, dividing training and test data, training a Random Forest model, and evaluating its performance using the confusion matrix, accuracy, precision, recall, and F1-score metrics using Python. The results are expected to provide an overview of Random Forest's performance in distinguishing Phishing URLs from legitimate URLs based on a combination of URL features and web page characteristics. These findings can also serve as a reference for developing Phishing detection systems as a form of technological support for Phishing prevention.

Keywords: *Phishing, Phishing URL, Random Forest, Machine learning, Python.*

KATA PENGANTAR

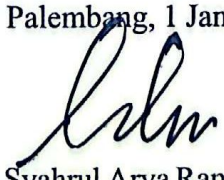
Puji syukur ke hadirat Allah SWT atas segala rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul 'Analisis Klasifikasi URL *Phishing* menggunakan *Random Forest* berdasarkan Fitur URL'. Penyusunan skripsi ini dimaksudkan untuk memenuhi salah satu syarat kelulusan pada Program Studi Teknologi Informasi, Universitas Muhammadiyah Palembang.

Penulis menyadari bahwa keberhasilan penyelesaian karya ilmiah ini tidak terlepas dari bantuan, bimbingan, serta dukungan moral dari berbagai pihak. Oleh karena itu, dengan kerendahan hati, penulis menyampaikan apresiasi dan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Abid Djazuli, S.E., M.M selaku Rektor Universitas Muhammadiyah Palembang.
2. Bapak Ir. A. Junaidi, M.T. selaku Dekan Fakultas Teknik Universitas Muhammadiyah Palembang.
3. Bapak Karnadi, S,Kom., M.Kom. selaku Ketua Program Studi Teknologi Informasi, serta selaku Dosen Pembinmbing Project Pra Tugas Akhir.
4. Bapak Kemas Muhammad Wahyu Hidayat, S.Kom., M.Kom selaku dosen pembmbing utama Skripsi penelitian
5. Bapak Dr. Lucky Indra Kesuma, S.SI., M.Kom selaku dosen pendamping Skripsi peneltian
6. Bapak / Ibu Dosen dan Staff Program Studi Teknologi Informasi

Penulis menyadari sepenuhnya bahwa skripsi ini masih memiliki berbagai kekurangan, baik dari aspek substansi maupun sistematika penulisan. Oleh sebab itu, berbagai kritik serta saran yang bersifat konstruktif sangat dinantikan demi perbaikan dan pengembangan karya ilmiah ini di masa yang akan datang

Palembang, 1 Januari 2026



Syahrul Arya Ramadhan

162022065

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1. Latar Belakang

Transformasi digital yang dipicu oleh pesatnya perkembangan teknologi informasi dan komunikasi dalam beberapa dekade terakhir telah merombak tatanan kehidupan secara fundamental. Peralihan aktivitas dari ranah konvensional menuju platform digital mencakup spektrum yang luas, mulai dari interaksi sosial melalui jejaring media, digitalisasi transaksi perbankan (e-banking), hingga implementasi layanan administrasi publik berbasis web. Meskipun integrasi internet menawarkan efisiensi dan aksesibilitas informasi yang masif, fenomena ini secara simultan melahirkan tantangan baru dalam domain keamanan siber (*cybersecurity*). Seiring dengan meningkatnya ketergantungan masyarakat terhadap ekosistem digital, kompleksitas ancaman siber pun berkembang menjadi risiko sistemik yang memerlukan perhatian serius.

Phishing merupakan salah satu serangan nerbahaya yang bersifat mutlak juga frekuensi kejadiannya kan meningkat. Sebagai bentuk serangan rekayasa sosial, *Phishing* berupaya mengeksploitasi informasi sensitif seperti kredensial akun, data pribadi, hingga detail finansial dengan memanipulasi identitas entitas tepercaya. Media penyebarannya sangat variatif, mencakup surel, pesan instan, serta platform media sosial yang mengarahkan korban ke situs web tiruan. Pada dasarnya, keberhasilan *Phishing* lebih dominan dipengaruhi oleh faktor manusia (human factor), yakni rendahnya literasi digital dan kewaspadaan pengguna, dibandingkan dengan celah pada sistem keamanan teknologi itu sendiri [1]

Di antara berbagai bentuk *Phishing*, serangan berbasis URL merupakan salah satu metode yang paling umum digunakan oleh pelaku. URL *Phishing* biasanya disisipkan dalam bentuk tautan yang tampak meyakinkan dan sering kali menyerupai domain resmi dari suatu layanan populer. Ketika pengguna mengklik tautan tersebut, pengguna akan diarahkan ke halaman palsu yang dirancang menyerupai situs asli. Tanpa disadari, pengguna kemudian memasukkan data pribadi atau kredensial akun, yang selanjutnya dapat dimanfaatkan oleh pelaku untuk tujuan kejahatan seperti pencurian identitas, pembobolan akun, hingga penipuan finansial. Seiring waktu, teknik penyamaran URL *Phishing* semakin berkembang, sehingga sulit dibedakan dengan URL asli hanya melalui pengamatan visual sederhana.

Sejumlah studi terdahulu telah mengukuhkan peran krusial *security awareness* sebagai garda terdepan dalam mitigasi serangan *Phishing*. Pengguna dengan literasi siber yang mumpuni secara teoretis memiliki kapabilitas untuk mengidentifikasi anomali, seperti struktur tautan yang mencurigakan, ketidakabsahan domain, serta pola komunikasi manipulatif yang bersifat urgensi [2]. Kendati demikian, restrukturisasi kesadaran pengguna secara mandiri belum menjadi solusi komprehensif dalam menanggulangi ancaman ini. Hal tersebut disebabkan oleh eskalasi kecanggihan taktik yang diterapkan oleh aktor ancaman, mencakup manipulasi domain, penggunaan *Top Level Domain* (TLD) yang menyerupai entitas resmi, hingga rekayasa konten web yang identik dengan situs orisinal [3]. Dinamika ini menyebabkan identifikasi manual kehilangan efektivitasnya, khususnya bagi pengguna awam yang memiliki keterbatasan

kompetensi teknis dalam memvalidasi integritas sebuah URL.

Sebagai upaya memitigasi keterbatasan faktor manusia, urgensi implementasi solusi teknologi melalui sistem deteksi otomatis berbasis *machine learning* menjadi kian krusial. Pendekatan ini beroperasi dengan mengekstraksi pola-pola laten dari data historis guna melakukan klasifikasi data baru dengan tingkat akurasi yang lebih presisi. Dalam penelitian ini, identifikasi URL *Phishing* dilakukan melalui analisis parameter multidimensi yang melampaui struktur fundamental URL. Analisis ini mengintegrasikan atribut domain, variasi *Top Level Domain* (TLD), serta protokol keamanan HTTPS, yang kemudian dikolaborasikan dengan metrik korelasi konten halaman web seperti *LineOfCode* dan *DomainTitleMatchScore* [4].

Konstruksi model klasifikasi dalam penelitian ini mengimplementasikan algoritma *Random Forest,* sebuah metode ensemble learning yang bekerja dengan mengintegrasikan sejumlah pohon keputusan (*decision trees*) untuk mengoptimalkan akurasi serta meminimalisir risiko overfitting [5]. Selain karakteristiknya yang stabil dalam menangani dataset multivariat, algoritma ini dipilih karena kapabilitasnya dalam menyajikan analisis feature importance, yang memungkinkan peneliti mengidentifikasi parameter paling signifikan dalam proses deteksi. Dataset yang menjadi basis penelitian ini diperoleh dari platform Kaggle melalui dataset '*Phishing URL Website*', yang memuat representasi URL berbasis fitur dengan pelabelan biner, yaitu *Phishing* (1) dan non-*Phishing* (0).

Dataset ini mengintegrasikan fitur berbasis struktur URL dengan karakteristik teknis halaman web, sehingga menyajikan dimensi analisis yang lebih

komprehensif dalam proses klasifikasi. Implementasi dataset ini memfasilitasi penelitian sistematis melalui pendekatan *supervised learning*, di mana model diinstruksikan menggunakan data yang telah terlabeli secara eksplisit. Sebelum fase pelatihan, data melalui tahapan *preprocessing* yang ketat, mencakup eliminasi redundansi (data *cleaning*), penanganan nilai yang hilang (*missing values*), serta transformasi fitur kategorikal melalui teknik encoding. Selanjutnya, model *Random Forest* dievaluasi menggunakan metode *train-test split* guna menjamin objektivitas hasil pengukuran performa model.

Evaluasi performa model dalam penelitian ini dikuantifikasi menggunakan serangkaian metrik klasifikasi yang komprehensif, meliputi *accuracy*, *precision*, *recall*, dan *F1-score*, yang diperkuat dengan analisis *confusion matrix*. Diversifikasi metrik ini bertujuan untuk memvalidasi efikasi model secara menyeluruh, khususnya dalam domain deteksi URL *Phishing*. Dalam konteks keamanan siber, metrik *recall* menjadi parameter kritis karena merepresentasikan kapabilitas model dalam mengidentifikasi entitas berbahaya secara maksimal. Hal ini krusial untuk memitigasi risiko false negative, di mana kegagalan model dalam mendeteksi URL *Phishing* dapat berakibat fatal karena dianggap sebagai tautan yang aman oleh pengguna.

Penelitian ini diproyeksikan memberikan kontribusi signifikan dalam bentuk analisis empiris terhadap performa algoritma *Random Forest* dalam mengklasifikasikan URL *Phishing* melalui integrasi fitur struktural URL dan karakteristik teknis halaman web. Melampaui pendekatan konvensional yang cenderung menitikberatkan pada aspek *security awareness*, studi ini

menghadirkan pelengkap strategis berupa solusi teknologi berbasis deteksi otomatis. Hasil penelitian ini diharapkan dapat menjadi rujukan fundamental bagi pengembangan sistem proteksi *Phishing* di masa depan.

1.2. Identifikasi Masalah

Dari pemaparan latar belakang di atas, maka bisa dirumuskan beberapa poin permasalahan sebagai berikut:

1. URL *Phishing* semakin kompleks, sehingga sulit diidentifikasi secara manual oleh pengguna.
2. Belum diketahui performa *Random Forest* dalam mengklasifikasikan URL *Phishing* berdasarkan fitur URL pada dataset yang digunakan..
3. Belum diketahui metrik evaluasi secara terukur (*accuracy*, *precision*, *recall*, *F1-score*) untuk memastikan kualitas model dalam mendeteksi URL *Phishing*.

1.3. Rumusan masalah

Bertolak dari identifikasi masalah yang telah dipaparkan, jadi rumusan masalah dalam penelitian ini ditetapkan sebagai berikut:

1. Bagaimana tahapan *preprocessing* dataset fitur URL agar dapat digunakan untuk proses klasifikasi URL *Phishing*?
2. Bagaimana penerapan *Random Forest* dalam mengklasifikasikan URL *Phishing* berdasarkan fitur URL?
3. Bagaimana hasil evaluasi performa *Random Forest* dalam mendeteksi URL *Phishing* berdasarkan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*?

1.4. Batasan Penelitian

Agar penelitian ini lebih terarah dan tidak menyimpang dari tujuan yang telah ditetapkan, maka diberikan beberapa batasan penelitian sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini merupakan dataset publik yang diperoleh dari platform Kaggle dengan nama *Phishing URL Website Dataset* yang diunggah oleh akun wendy0701.
2. Ruang lingkup penelitian ini difokuskan pada penggunaan delapan fitur utama sebagai parameter analisis, yaitu *Domain*, *Top Level Domain (TLD)*, *NoOfOtherSpecialCharsInURL*, *HTTPS*, *LineOfCode*, *Title*, *DomainTitleMatchScore*, dan *URLTitleMatchScore*.
3. Penelitian ini menggunakan pendekatan *supervised learning* dengan target klasifikasi berupa label *Phishing* (1) dan *legitimate* (0).
4. Algoritma yang digunakan dalam penelitian ini hanya terbatas pada *Random Forest* tanpa melakukan perbandingan dengan algoritma lain.
5. Tahapan pengolahan data serta pengembangan model diimplementasikan menggunakan bahasa pemrograman Python dengan mengintegrasikan berbagai pustaka (*library*) khusus *machine learning.*
6. Kinerja model diukur melalui serangkaian metrik evaluasi klasifikasi yang komprehensif, meliputi *confusion matrix*, tingkat akurasi *(accuracy)*, presisi *(precision)*, *recall*, serta *F1-score.*

1.5. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disusun, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis tahapan *preprocessing* dataset URL *Phishing* agar data siap digunakan dalam proses klasifikasi.
2. Menerapkan algoritma *Random Forest* dalam mengklasifikasikan URL *Phishing* berdasarkan fitur URL dan karakteristik halaman web.
3. Menganalisis efektivitas model *Random Forest* dalam mengidentifikasi URL *Phishing* dengan mengacu pada metrik evaluasi yang terdiri dari *confusion matrix, accuracy, precision, recall, serta F1-score.*

1.6. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis, sebagai berikut:

1.6.1. Manfaat Teoritis

Pada penelitian berikut berharap bisa menghasilkan kontribusi dalam perkembangan ilmu pengetahuan pada bidang keamanan siber (*cybersecurity*) dan *machine learning*, khususnya dalam analisis klasifikasi URL *Phishing* berbasis fitur URL. Di samping itu, diharapkan hasil ini dapat membantu kontribusi teoretis juga jadi rujukan bagi studi di masa depan.

1.6.2. Manfaat Praktis

1. Bagi peneliti, studi ini diharapkan dapat memperdalam pemahaman mengenai implementasi algoritma *Random Forest* serta memberikan wawasan baru terkait efektivitas fitur-fitur teknis dalam mengidentifikasi ancaman *Phishing*.
2. Bagi akademisi, penelitian ini dapat digunakan sebagai bahan referensi dalam pembelajaran terkait *machine learning* dan keamanan siber.
3. Bagi pengembang sistem, hasil penelitian ini dapat dijadikan dasar dalam pengembangan sistem deteksi *Phishing* secara otomatis.
4. Bagi pengguna internet, penelitian ini diharapkan dapat meningkatkan perlindungan terhadap ancaman *Phishing* melalui pemanfaatan teknologi deteksi otomatis.

1.7. Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun untuk memberikan gambaran umum mengenai isi setiap bab dalam skripsi. Adapun sistematika penulisan adalah sebagai berikut:

BAB I PENDAHULUAN

Isi dari bab ini yaitu latar belakang, identifikasi masalah, rumusan masalah, batasan penelitian, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Isi dari bab ini yaitu landasan teori yang mendukung penelitian, kajian penelitian terdahulu (*state of the art*), serta kerangka konseptual penelitian.

BAB III METODE PENELITIAN

Isi dari bab ini yaitu menguraikan dengan detail mengenai desain dan pendekatan penelitian, sumber serta teknik pengumpulan data, metodologi yang diterapkan, kerangka konseptual, prosedur penelitian, teknik analisis data, hingga lini masa pelaksanaan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Isi dari bab ini yaitu hasil pengolahan data, implementasi model *Random Forest*, evaluasi performa model, serta analisis dan pembahasan hasil penelitian.

BAB V PENUTUP

Isi dari bab ini yaitu kesimpulan dari hasil penelitian serta saran yang bisa digunakan untuk penelitian selanjutnya.

DAFTAR PUSTAKA

[1] D. W. Adani, I. Nurhayati, dan R. E. Mirati, "The Importance of *Security awareness* in Combating *Phishing* Attacks: A Case Study of Bank Rakyat Indonesia," *SOSHUM: Jurnal Sosial dan Humaniora*, vol. 14, no. 3, 2024.

[2] A. N. Khomarudin, R. Aulia, J. Jamaluddin, R. Novita, M. Ikhsan, M. N. Afif, dan H. Suherlan, "Penyuluhan Internet Sehat tentang Bahaya Phising dan Dampak Negatifnya bagi Siswa SMK N 4 Payakumbuh," *Interaksi: Jurnal Pengabdian Kepada Masyarakat*, vol. 2, no. 1, pp. 16–22, Jun. 2025.

[3] M. N. H. Jatsono, F. F. Suri, R. D. Rohmaniar, M. A. Al-Fatih, V. H. Putra, D. W. Falih, S. A. Kristyan, dan R. Yasirandi, "Masyarakat Cakap *Digital*: Sosialisasi Bahaya dan Pencegahan Serangan *Phishing*," *Jurnal Pemanfaatan Teknologi untuk Masyarakat (JAPATUM)*, vol. 3, no. 4, pp. 19–26, Sep. 2023, doi: 10.59328/JAPATUM.2023.2.3.65.

[4] I. D. Sari, D. Hariyadi, R. Sahtyawan, dan N. I. Kusumaningtyas, "Analisis Tingkat *Security awareness*–Personal Threat terhadap Ancaman *Phishing* dengan Metode Technology Threat Avoidance Theory (TTAT)," *Teknomatika: Jurnal Informatika dan Komputer*, vol. 16, no. 2, pp. 49–55, Sep. 2023, doi: 10.30989/teknomatika.v16i2.1250.

[5] K. Z. Ansyafa, M. Fajarudin, M. Fadhil, dan S. N. Neyman, "Analisis Keamanan Media Sosial terhadap Serangan Phising *Online* menggunakan Metode Zphisher dan *Social engineering* Toolkit," *Journal of Internet and Software Engineering*, vol. 1, no. 4, pp. 1–10, Okt. 2024, doi: 10.47134/pjise.v1i4.2641.

[6] S. Pohan, D. Irfan, Y. I. M. Hasibuan, dan I. Cahyani, "Edukasi dan Simulasi Deteksi Serangan *Phishing* pada *Email* Akademik Mahasiswa Berbasis *Social engineering*," *Jurnal Abdimas Ika Bina*, vol. 1, no. 1, pp. 1–8, Sep. 2023.

[7] A. D. Prasetyo, H. B. Seta, dan I. W. Widi P., "Analisis *Digital* Forensik Spear *Phishing* Menggunakan Metode National Institute of Justice (Studi Kasus: Instagram Verified Account)," *Jurnal Informatik*, vol. 19, no. 1, pp. 58–67, Apr. 2023.

[8] D. A. Akbar, M. R. E. Kurnia, R. M. G. S. Bintang, dan R. Purwoko, "Analisis Web *Phishing* Menggunakan Metode OSCAR Forensic (Studi Kasus: Follower Instagram Gratis)," *Jurnal Teknik Informatika*, vol. 3, no. 1, pp. 18–24, Feb. 2024.

[9] K. Isadora, N. P. Aqila, H. Gustina, A. Nabila, dan Nurfitriana, "Analisis Modus Phising terhadap Whatsapp," *Jurnal Akuntansi, Bisnis dan Ekonomi Indonesia*, vol. 3, no. 2, pp. 45–52, Agu. 2024.

[10] R. K. Sujiwana, A. F. A. Ridho, D. C. Aryanti, dan N. A. Rakhmawati, "Analisis Bibliometrik Mengenai Serangan *Phishing* dan WhatsApp Menggunakan VOSviewer," *Jurnal Esensi Infokom*, vol. 8, no. 1, pp. 101–105, Mei 2024.

[11] I. L. Putra, F. Siahaan, M. I. Raditya, M. O. A. Purba, dan I. Gunawan, "Analisis Keamanan Sistem Operasi Android terhadap Serangan *Phishing* pada Aplikasi E-Wallet," *Jurnal Inovasi Artificial Intelligence & Komputasional Nusantara (JIKOMNUS)*, vol. 2, no. 1, pp. 22–26, Jul. 2025.

[12] I. A. Saputro, L. Sugiarto, dan F. S. Nugraha, "Analisis Kesadaran Masyarakat terhadap Bahaya Internet *Phishing* Menggunakan K-Means Clustering," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 9, no. 2, pp. 139–146, Des. 2024.

[13] A. Nofiyan dan Mushlihudin, "Analisis Forensik pada Web *Phishing* Menggunakan Metode National Institute of Standards and Technology (NIST)," *Jurnal Sarjana Teknik Informatika*, vol. 8, no. 2, pp. 11–23, Jun. 2020.

[14] K. Saidi dan Y. Prayudi, "Analisis Indikator Utama dalam Information Security–Personality Threat terhadap *Phishing* Attack Menggunakan Metode Technology Threat Avoidance Theory (TTAT)," *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, vol. 6, no. 1, pp. 21–30, Feb. 2021.

[15] A. A. Ardelia, Q. A. R. 'Aisy, dan Santikasari, "Analisis Keamanan dan Privasi Data Instagram terhadap Ancaman *Phishing* di Era *Digital*," dalam *Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, Surakarta, Indonesia, Jul. 2024, pp. 366–371.

[16] D. Bramasta, R. R. Ardianto, dan N. S. S. Dewi, "Analisis Strategi Efektif dalam Mendeteksi dan Mencegah Serangan *Phishing* Melalui Undangan Elektronik Berformat .Apk," dalam *Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, Surakarta, Indonesia, Jul. 2024, pp. 436–441.

[17] T. F. Ramadhan, I. Ramadhan, dan A. A. Pangestu, "Analisis Keamanan Teknologi dalam Menghadapi Ancaman *Phishing*," *Jurnal Teknik Informatika*, pp. 1–6, 2024.

[18] N. Vadila dan A. R. Pratama, "Analisis Kesadaran Keamanan terhadap Ancaman *Phishing*," Universitas Islam Indonesia, Yogyakarta, 2020.

[19] M. Irdi, R. Aditya, A. Ramdhani, D. D. Nugraha, M. F. A. Febrian, dan Nugraha, "Analisis Masalah Serangan *Phishing* pada Penggunaan *Email*," dalam *Seminar Nasional Teknologi Informasi, Mekatronika dan Ilmu Komputer (SENTIMETER)*, Universitas Nusa Putra, Sukabumi, Indonesia, Mei 17, 2025, pp. 1–5.

[20] L. Wijaya dan E. K. Nurnawati, "Analisis Kesadaran Mahasiswa Yogyakarta tentang *Phishing* pada *Online* Banking," *Jurnal Dinamika Informatika*, vol. 11, no. 2, pp. 113–122, Sep. 2022.

[21] S. Nailah dan R. Rosnelly, "Analisis Ancaman *Phishing* di *Platform* Media Sosial *Email* Menggunakan Metode *Digital* Forensic Research Workshop," *Jurnal Rekayasa Sistem*, vol. 3, no. 2, pp. 450–461, Mei 2025.

[22] M. A. Al Fadillah, M. G. Ramadhan, dan M. E. Ariefiandi, "Analisis Ancaman *Phishing* terhadap Penggunaan *E-commerce* di Indonesia," *Journal of Information and Information Security (JIFORTY)*, vol. 5, no. 2, pp. 85–96, Des. 2024.

[23] D. K. Lahagu, D. Zalukhu, F. M. N. Hura, F. Harefa, P. B. Telaumbanua, dan O.

Laia, “Analisis Potensi Kerentanan Terhadap Serangan *Phishing* pada Website sttsundermann.siakadcloud.com Menggunakan Simulasi Lingkungan Kali Linux dan Ngrok,” *Jurnal Komputer Teknologi Informasi Sistem Komputer (JUKTISI)*, vol. 4, no. 2, pp. 390–397, Sep. 2025, doi: 10.62712/juktisi.v4i2.398.

[24] O. Iskandar, “Analisis Kejahatan *Online Phishing* pada Masyarakat,” *Jurnal Fakultas Hukum Universitas Bhayangkara Jakarta Raya*, vol. 1, no. 2, pp. 32–36, Jun. 2024.

[25] N. B. Putri dan A. W. Wijayanto, “Analisis Komparasi Algoritma Klasifikasi Data Mining dalam Klasifikasi Website *Phishing*,” *Komputika: Jurnal Sistem Komputer*, vol. 11, no. 1, pp. 59–66, Apr. 2022.

[26] Z. Alfharizi, L. I. Kesuma, dan D. Haryanto, “Analysis of critical success factors in information technology project management: A literature review,” Journal of Artificial Intelligence and Engineering Applications, vol. 5, no. 2, Feb. 2026, e-ISSN: 2808-4519.

[27] A. R. Dwi, D. Haryanto, dan Apriansyah, “Analysis of LAN network quality at PTPN7 Senabing Unit in Lahat Regency using QoS (Quality of Service),” Jurnal Teknologi dan Open Source, vol. 8, no. 2, pp. 883–890, Dec. 2025, doi: 10.36378/jtos.v8i2.5030.

[28] F. Agustina, D. Haryanto, dan M. Ihsan, “Analysis and optimization of the LAN network using the load balancing technique (Case Study: University of Muhammadiyah Palembang Campus),” Jurnal Teknologi dan Open Source, vol. 8, no. 2, pp. 1165–1172, Dec. 2025, doi: 10.36378/jtos.v8i2.5166.

[29] wendy0701, “*Phishing URL Website* Dataset,” Kaggle. [Online]. Available: https://www.kaggle.com/datasets/wendy0701/phising

[30] T. D. Sitompul, D. P. Ananta, M. R. Hanif, M. Wati, dan Haviluddin, “Penerapan K-Means Clustering dalam Analisis URL *Phishing* untuk Identifikasi Risiko Keamanan Menggunakan Model PCA,” Adopsi Teknologi dan Sistem Informasi *(ATASI)*, vol. 4, no. 2, pp. 127–136, 2025.

[31] I. Radiansyah, Candiwan, dan Y. Priyadi, “Analisis Ancaman *Phishing* dalam Layanan Online Banking,” Ekonomika-Bisnis, vol. 7, no. 1, pp. 1–14, 2016.

[32] Mushlihudin dan A. Nofiyan, “Analisis Forensik pada Web *Phishing* Menggunakan Metode *National Institute of Standards and Technology*,” *CYBERNETICS*, vol. 4, no. 2, pp. 79–92, 2020.