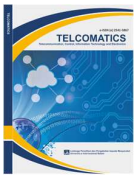
Available online at journal.uib.ac.id/index.php/telcomatics

TELCOMATICS

ISSN (online) 2541-5867



Pendekatan Klasifikasi Random Forest untuk Identifikasi URL Berbahaya yang Akurat

Elvert[1], Haeruddin[2*], Andik Yulianto[3], Sabariman[4]

[1,2]Prodi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Internasional Batam, Jalan Gajah Mada, Baloi Sei Ladi, Batam, Indonesia

[3,4]Prodi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Internasional Batam, Jalan Gajah Mada, Baloi Sei Ladi, Batam, Indonesia

ARTICLE INFO

Keywords:
Machine Learning, Random Forest, Security

Received: August 6, 2025
Revised: November 2, 2025
Accepted: December 10, 2025

***Corresponding author:**
E-mail: haeruddin@uib.ac.id

DOI: 10.37253/telcomatics.v10i2.11173

ABSTRACT

Internet users currently face significant risks from malicious URLs that facilitate phishing attacks, malware distribution, and data theft. Traditional blacklisting methods have become ineffective against evolving cyberattack techniques. This study proposes a Random Forest classification approach for more accurate malicious URL detection, focusing on critical URL features including URL length, presence of special keywords, subdomain structure, and special character usage. These features train the Random Forest model to distinguish between safe and malicious URLs. We evaluate model effectiveness using accuracy, precision, and recall metrics. This research aims to develop a Random Forest-based malicious URL detection system that is more accurate and adaptive than conventional methods. The study examines both the advantages and limitations of this approach, along with its potential as a reliable detection solution for dynamic digital environments. Evaluation results demonstrate an overall accuracy of 94%, weighted average F1-score of 0.94, and macro average F1-score of 0.94.

I. PENDAHULUAN

Dalam dekade terakhir, internet telah menjadi elemen penting dalam kehidupan manusia sehingga sulit dibayangkan bagaimana dunia berfungsi tanpanya. Menurut laporan yang dirilis pada Januari 2021, terdapat 4,66 miliar individu yang aktif menggunakan internet di seluruh dunia. Angka ini mewakili sekitar 59,5% dari total populasi dunia. Selain itu, sekitar 92,6% pengguna internet mengaksesnya melalui perangkat smartphone [1].

Kebanyakan orang menggunakan World Wide Web (WWW) untuk berbagai aktivitas sehari-hari seperti perbankan online, e-commerce, dan berbagi informasi. Namun, ada juga segmen web yang dirancang untuk melakukan aktivitas ilegal seperti mencuri data penting. Upaya pencurian kredensial pengguna ini dikenal sebagai 'serangan phishing', dan tautan web dengan niat jahat disebut sebagai Uniform Resource Locators (URL) berbahaya.

Phishing adalah serangan rekayasa sosial yang diidentifikasi sebagai metode paling umum digunakan oleh penjahat siber untuk mendapatkan akses ke informasi pribadi pengguna seperti detail kartu kredit, nama pengguna, dan kata sandi [2], [3] Phishing telah menjadi taktik favorit penjahat siber karena biayanya rendah dan tidak memerlukan keahlian teknis yang mendalam. Sebagian besar upaya phishing dimulai dengan email spam yang sering berisi tautan ke situs web penipuan. Pada April 2020, layanan email besar melaporkan memblokir lebih dari 100 juta email spam setiap hari, termasuk 18 juta yang terkait dengan penipuan phishing COVID-19 [4]. Serangan phishing semakin menjadi masalah besar bagi individu dan organisasi dalam beberapa tahun terakhir. Laporan dari Anti-Phishing Working Group (APWG) menemukan bahwa serangan phishing meningkat dua kali lipat dari awal 2020 hingga kuartal ketiga 2021 [5].

Deteksi phishing pada halaman web menggunakan pembelajaran mesin umumnya melibatkan ekstraksi fitur relevan dari berbagai komponen halaman web dan pelatihan model prediksi berbasis pembelajaran mesin menggunakan fitur-fitur ini. Penelitian komprehensif telah dilakukan pada berbagai jenis fitur, termasuk fitur leksikal, fitur berbasis host, fitur berbasis konten, serta fitur berdasarkan konteks dan popularitas [4].

Beberapa penelitian telah menyelidiki berbagai metode untuk meningkatkan deteksi URL phishing dengan hasil yang cukup sukses, termasuk Temporal Convolutional Network (TCN), model regresi logistik, naive bayes, SVR, kNN, dan SVM.[5] Studi sebelumnya menggunakan TCN menunjukkan efektivitas deep learning dalam mendeteksi URL phishing dengan akurasi tinggi melalui embedding semantik URL. Namun, pendekatan ini memiliki kelemahan dalam hal inefisiensi komputasi, kurangnya interpretabilitas, dan ketergantungan pada dataset berskala besar. Untuk mengatasi keterbatasan tersebut, kami mengusulkan klasifier berbasis Random Forest yang menyeimbangkan akurasi, kecepatan, dan kemudahan interpretasi, sekaligus tetap tangguh terhadap strategi phishing yang terus berevolusi.

Random Forest adalah pengklasifikasi yang terdiri dari kumpulan pengklasifikasi berbasis pohon di mana setiap pohon memberikan voting untuk kelas paling populer pada input tertentu [6]. Random Forest biasanya lebih unggul daripada regresi linier dalam tugas prediksi karena mampu beradaptasi dengan pola nonlinier dalam data [7].

Berdasarkan analisis APWG [5], teknik phishing terus berevolusi dengan 72% serangan baru menggunakan domain yang sebelumnya tidak dikenal. Kondisi ini membutuhkan pendekatan deteksi yang tidak hanya akurat tetapi juga efisien dan mudah diinterpretasikan. Meskipun solusi berbasis deep learning seperti TCN [5] mencapai akurasi tinggi, implementasinya menghadapi kendala dalam hal kebutuhan komputasi dan interpretabilitas. Random Forest dipilih dalam penelitian ini karena kemampuannya dalam menangani data tidak seimbang dan menyediakan analisis pentingnya fitur, sekaligus mempertahankan efisiensi komputasi yang dibuktikan dalam penelitian terkait [8]. Berdasarkan teori Breiman (2001), ensemble methods seperti Random Forest mampu mengurangi overfitting tanpa mengorbankan generalisasi.[6] Dengan memanfaatkan non-linearitas alami dari pohon keputusan, Random Forest classifier sangat cocok untuk menangkap pola dan hubungan kompleks dalam dataset, menjadikannya alat yang ampuh untuk mengidentifikasi berbagai jenis URL berbahaya seperti *phishing*, *malware*, dan *defacement*.

II. TINJAUAN PUSTAKA

A. Penelitian Sebelumnya

Menurut Safi A dan Singh S dalam penelitiannya yang berjudul "A systematic literature review on phishing website detection techniques" menganalisis 128 penelitian dan menyimpulkan bahwa algoritma Random Forest memperoleh 92-95% F1-score dalam mendeteksi phishing, mengungguli SVM dan regresi logistik karena ketangguhannya dengan fitur hibrida (berbasis leksikal + host)[2].

Menurut Opara C. , Chen Y., dan Wei B. dalam penelitian mereka yang berjudul "Look before you leap: Detecting phishing web pages by exploiting raw URL and HTML characteristics" yang menggabungkan fitur URL (seperti panjang dan karakter khusus) dengan analisis HTML untuk mencapai akurasi 98,2% menggunakan Random Forest [4].

Beberapa pendekatan deep learning seperti Temporal Convolutional Network (TCN) telah mencapai akurasi tinggi dalam deteksi URL phishing dengan menganalisis pola temporal pada string URL yang dilakukan oleh Remmide et al. dalam penelitian mereka yang berjudul "Detection of Phishing URLs Using Temporal Convolutional Network" [5].

Dalam Penelitian Vajrobov et al. Yang berjudul "Mutual information based logistic refression for phishing URL detection" mereka menggunakan regresi logistik yang dioptimalkan dengan informasi timbal balik mencapai akurasi 94,1% [9].

Menurut Banik B dan Sarma A dalam penelitian mereka yang berjudul "Phishing Url Deteciton Using LSTM based Ensemble Learning Approaches" LSTM ensemble mencapai akurasi 99,65%, namun membutuhkan sumber daya komputasi tinggi. [10]

Menurut penelitian Sheikhi dan Kostakos, PSO-XGBoost mencapai akurasi 98,42% tetapi bergantung pada komputasi yang intensif dan mahal. Penelitian ini menggunakan algoritma Firefly untuk memilih 23 fitur leksikal atau host. Algoritma ini mendapatkan hasil yang akurat tetapi waktu komputasi menjadi lebih lama [11].

B. Phishing

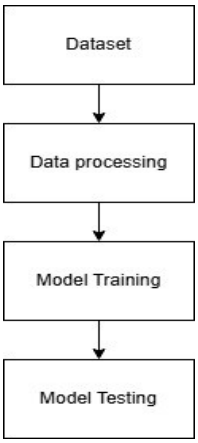
Phishing adalah serangan rekayasa sosial yang diidentifikasi sebagai metode paling umum yang digunakan oleh penjahat siber untuk mendapatkan akses ke informasi pribadi pengguna internet seperti informasi kartu kredit, nama pengguna, dan kata sandi. Serangan phishing memiliki banyak bentuk dan biasanya melibatkan berbagai saluran komunikasi, seperti email, pesan singkat, kode Quick Response (QR), dan media sosial. Penyerang biasanya meniru bank terkenal, agen kartu kredit, atau situs web *e-commerce* terkenal untuk mengintimidasi atau meyakinkan pengguna agar masuk ke situs web phishing dan memberikan informasi penting[2], [12], [13].

C. Algoritma Random Forest

Random Forest adalah algoritma machine learning yang menggabungkan beberapa decision tree untuk membuat prediksi, meningkatkan akurasi dan stabilitas dibandingkan dengan pohon keputusan tunggal. Random Forest menggunakan model prediksi berupa diagram pohon dengan node yang merepresentasikan fitur/variabel, cabang yang mewakili aturan pemilahan, dan leaf/simpul terminal sebagai kelas prediksi atau keputusan. [6]

D. URL

URL adalah singkatan dari Uniform Resource Locator, yang merupakan alamat global dokumen dan sumber daya tertentu di internet, seperti situs web, halaman web, atau file. Sebuah URL terdiri dari beberapa komponen seperti protokol yang biasanya digunakan http atau https dan nama domain yang berfungsi sebagai identifikasi untuk website seperti google.com [14], [15].



Gambar 1. Tahapan Penelitian

III. METODE PENELITIAN

A. Tahapan Penelitian

Penelitian ini terdiri dari beberapa tahapan seperti pada Gambar 1. Dataset yang digunakan dalam penelitian ini merupakan dataset diambil dari Kaggle. Dataset ini berisi URL yang diklasifikasikan menjadi 4 kategori yaitu, *Benign, Defacement*, *Phishing* dan *Malware*. Dataset yang digunakan terdiri dari 651191 data yang mana mencakup 428103 link benign, 96457 link defacement, 94111 link phishing, dan malware 32530 link.

Pada tahap pengolahan data melibatkan pembersihan data, penyeimbangan data dan tokenisasi fitur leksikal dari url. Proses penyeimbangan data dilakukan undersampling sehingga dataset terdistribusi merata. Dataset juga dibagi menjadi 2 data yaittu training dan testing yang mana dataset training mencakup 80% dan testing 20%. Tujuan dari pengolahan data ini adalah untuk memastikan bahwa data yang digunakan dalam model adalah akurat dan siap digunakan untuk *model training*.

Pada tahap *Model Training* dataset yang telah diproses dibagi menjadi 2 yaitu dataset training dan testing dan digunakan dengan algoritma Random Forest. Model dilatih dengan menidentifikasi fitur leksikal dari url. Selanjutnya pada tahap *Model Testing* dilakukan pengujian pada model yang telah dilatih untuk melihat apakah model sudah layak atau belum.

IV. HASIL DAN PEMBAHASAN

Sumber data yang dipakai adalah data yang berisi kumpulan link malicious dan benign yang mana dataset diperoleh dari kaggle. Dataset juga dikategorikan menjadi 4 kelas yaitu phishing, malware, defacement, dan benign seperti pada Tabel 1.

Tabel 1. Spesifikasi Dataset

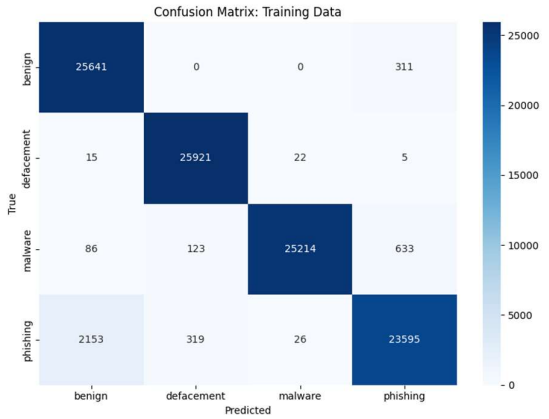
No.	Kelas	Jumlah link	Persentase
1.	Benign	428103	65,74%
2.	Defacement	96457	14,81%
3.	Phishing	94111	14,45%
4.	Malware	32520	5%

Dataset kemudian diseimbangkan untuk mengatasi masalah ketidakseimbangan kelas. Setelah proses undersampling, jumlah data pada setiap kelas menjadi sama, yaitu seperti data pada kelas Malware sebanyak 32.520 link. Tabel 2 menunjukkan komposisi dataset setelah diseimbangkan.

Tabel 2. Dataset setelah diseimbangkan

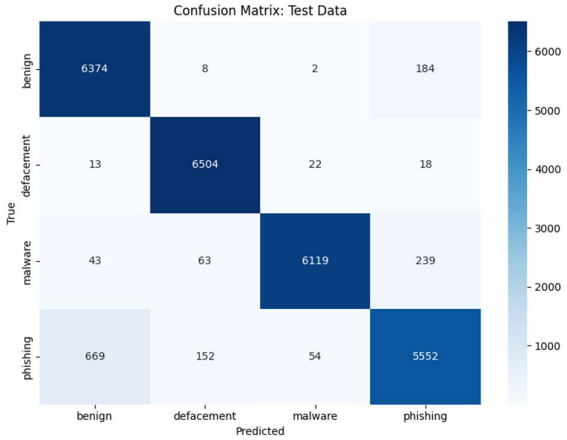
No.	Kelas	Jumlah link	Persentase
1.	Benign	428103	65,74%
2.	Defacement	96457	14,81%
3.	Phishing	94111	14,45%
4.	Malware	32520	5%

Setelah dataset diseimbangkan dengan metode undersampling, model dilatih dengan menggunakan fitur-fitur leksikal yang diekstraksi dari setiap url. Pelatihan model menghasilkan *confusion matrix* yang dapat dilihat pada Gambar 2.



Gambar 2. Confusion Matrix Pelatihan Model

Setelah dilatih, dilakukan pengujian terhadap model yang sudah dilatih dengan dataset sebelumnya. Tahap pengujian dilakukan untuk mengevaluasi kinerja model terhadap data yang belum pernah digunakan sebelumnya. Kinerja model dievaluasi menggunakan metrik Accuracy, Precision, Recall dan F1-Score.



Gambar 3. Confusion Matrix Pengujian Model

Gambar 3 menunjukkan hasil dari confusion amtrix setelah model diuji dengan dataset pengujian, model berhasil memprediksi url secara benar dengan data berikut yang dapat di lihat pada Tabel 3.

Tabel 3. Hasil evaluasi model

Kelas	Precision	Recall	F1- Score
Benign	0.90	0.97	0.93
Defacement	0.97	0.99	0.90
Phishing	0.99	0.95	0.97
Malware	0.93	0.86	0.89
Accuracy			0.94
Macro avg	0.94	0.94	0.94
Weighted avg	0.94	0.94	0.94

Secara keseluruhan, model yang kami kembangkan mampu mencapai tingkat akurasi sebesar 94%. Nilai Macro Average F1-Score dan Weighted Average F1-Score yang kami peroleh juga berada di angka 0.94. Kesamaan antara kedua nilai rata-rata ini menunjukkan bahwa kinerja model cukup konsisten di semua kelas, yang merupakan hasil yang diharapkan mengingat dataset pengujian telah diseimbangkan melalui teknik *undersampling*.

Seperti yang ditunjukkan dalam Tabel 3, kelas Defacement memiliki nilai Recall tertinggi, yaitu 0.99, yang menunjukkan kemampuan model yang sangat baik dalam mengenali URL jenis ini. Sebaliknya, kelas Malware menunjukkan nilai Recall terendah, yaitu 0.86, meskipun memiliki Precision tinggi sebesar 0.93. Ini menunjukkan bahwa prediksi model terhadap URL Malware cukup akurat, namun sekitar 14% dari URL Malware yang sebenarnya tidak berhasil terdeteksi.

Model juga menunjukkan kinerja yang baik pada kelas Phishing dan Benign, dengan nilai F1-Score masing-masing sebesar 0.97 dan 0.93. Hasil ini mencerminkan efektivitas fitur leksikal yang digunakan dalam mengidentifikasi pola-pola umum dari serangan tersebut. Secara keseluruhan, temuan ini mengonfirmasi bahwa model Random Forest yang diusulkan, dengan dukungan pra-pemrosesan data yang tepat, mampu menjadi solusi yang andal dan seimbang untuk deteksi URL berbahaya dalam skenario multi-kelas.

V. KESIMPULAN

Berdasarkan penelitian yang dilakukan, pendekatan Random Forest terbukti mampu menjadi solusi dalam membedakan URL yang aman dan yang berpotensi berbahaya, seperti phishing, malware, dan defacement. Hal ini dicapai melalui pemanfaatan fitur-fitur leksikal dari URL yang dapat merepresentasikan pola-pola yang sering muncul pada URL berbahaya. Kontribusi utama dari penelitian ini adalah memberikan model klasifikasi yang relatif sederhana, tidak membutuhkan komputasi tinggi, namun tetap memiliki performa yang cukup baik. Pendekatan ini juga memberikan gambaran bahwa model Random Forest bisa diimplementasikan dalam sistem deteksi keamanan yang membutuhkan waktu respons cepat serta hasil yang bisa dipahami dengan mudah. Namun, penelitian ini memiliki beberapa keterbatasan, di antaranya adalah keterbatasan jumlah fitur yang digunakan serta teknik undersampling yang bisa saja membuat model kehilangan sebagian informasi penting dari data yang awalnya tidak seimbang. Selain itu, model ini belum diuji pada data real-time atau lingkungan dunia nyata, sehingga validitasnya di luar data pelatihan masih perlu ditingkatkan. Penelitian selanjutnya disarankan untuk mengeksplorasi lebih banyak jenis fitur, menggabungkan Random Forest dengan algoritma lain, serta menguji model dalam sistem deteksi yang berjalan langsung agar hasilnya lebih aplikatif dan adaptif terhadap serangan siber yang terus berkembang

VI. DAFTAR PUSTAKA

[1] L. Tang and Q. H. Mahmoud, "A Survey of Machine Learning-Based Solutions for Phishing Website Detection," Sep. 01, 2021, *MDPI*. doi: 10.3390/make3030034.

[2] A. Safi and S. Singh, "A systematic literature review on phishing website detection techniques," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 590–611, Feb. 2023, doi: 10.1016/j.jksuci.2023.01.004.

[3] B. K. Gontla, P. Gundu, P. J. Uppalapati, K. Narasimharao, and S. M. Hussain, "A Machine Learning Approach to Identify Phishing Websites: A Comparative Study of Classification Models and Ensemble Learning Techniques," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 10, no. 5, pp. 1–9, 2023, doi: 10.4108/eetsis.vi.3300.

[4] C. Opara, Y. Chen, and B. Wei, "Look before you leap: Detecting phishing web pages by exploiting raw URL and HTML characteristics," *Expert Syst Appl*, vol. 236, Feb. 2024, doi: 10.1016/j.eswa.2023.121183.

[5] M. A. Remmide, F. Boumahdi, N. Boustia, C. L. Feknous, and R. Della, "Detection of Phishing URLs Using Temporal Convolutional Network," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 74–82. doi: 10.1016/j.procs.2022.10.209.

[6] Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

[7] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *Stata Journal*, vol. 20, no. 1, pp. 3–29, Mar. 2020, doi: 10.1177/1536867X20909688.

[8] Abdulhamit Subasi, Esraa Molah, Fatin Almkallawi, and Touseef J. Chaudhery, *Intelligent Phishing Website Detection using Random Forest Classifier*. IEEE, 2018.

[9] V. Vajrobol, B. B. Gupta, and A. Gaurav, "Mutual information based logistic regression for phishing URL detection," *Cyber Security and Applications*, vol. 2, Jan. 2024, doi: 10.1016/j.csa.2024.100044.

[10] B. Banik and A. Sarma, "PHISHING URL DETECTION USING LSTM BASED ENSEMBLE LEARNING APPROACHES," *International Journal of Computer Networks and Communications*, vol. 15, no. 1, pp. 17–33, Jan. 2023, doi: 10.5121/ijcnc.2023.15102.

[11] S. Sheikhi and P. Kostakos, "Safeguarding cyberspace: Enhancing malicious website detection with PSO[sbnd]optimized XGBoost and firefly-based feature selection," *Comput Secur*, vol. 142, Jul. 2024, doi: 10.1016/j.cose.2024.103885.

[12] B. Banik and A. Sarma, "Lexical Feature Based Feature Selection and Phishing URL Classification Using Machine Learning Techniques," in *Communications in Computer and Information Science*, Springer, 2020, pp. 93–105. doi: 10.1007/978-981-15-6318-8_9.

[13] R. Verma and A. Das, "What's in a URL: Fast feature extraction and malicious URL detection," in *IWSPA 2017 - Proceedings of the 3rd ACM International Workshop on Security and Privacy Analytics, co-located with CODASPY 2017*, Association for Computing Machinery, Inc, Mar. 2017, pp. 55–63. doi: 10.1145/3041008.3041016.

[14] D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL Detection using Machine Learning: A Survey," vol. 2019-August, doi: 10.48550/arXiv.1701.07179

[15] S. Mohanty and A. A. Acharya, "MFBFST: Building a stable ensemble learning model using multivariate filter-based feature selection technique for detection of suspicious URL," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 1668–1681. doi: 10.1016/j.procs.2023.01.145.