Klasifikasi URL Berbahaya Menggunakan Algoritma Random Forest Berbasis Fitur Struktural

I Gede Putra Wiratama[a1], Anak Agung Istri Ngurah Eka Karyawati[a2]

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]wiratama.2308561004@student.unud.ac.id
[2]eka.karyawati@unud.ac.id

Abstract

Phishing attacks remain a critical threat in the digital era, often exploiting deceptive URLs to trick users into divulging sensitive personal information. To address this issue, this study proposes a machine learning-based detection system using the Random Forest algorithm to identify phishing URLs based on structural features. The main objective of this research is to build an efficient and lightweight model that can detect phishing attempts in real-time without relying on third-party databases or content-based analysis. From the dataset used, 10 structural features were selected based on relevance and efficiency, such as the presence of IP addresses, use of HTTPS, domain age, and URL length. The model was trained and tested on a labeled dataset and evaluated using accuracy, confusion matrix, and classification report. The Random Forest model achieved a testing accuracy of 92.72%, with strong precision and recall values for both phishing and legitimate classes. The results indicate that the proposed approach is effective in distinguishing malicious URLs using only structural characteristics, making it a practical solution for enhancing cybersecurity at the URL level.

Keywords: *Phishing URL Detection, Random Forest, Information, Classification, URL-Based Feature Selection, Cybersecurity*

1. Pendahuluan

Perkembangan teknologi yang sangat pesat telah memainkan peran utama dalam meningkatnya popularitas dan pemasaran organisasi dalam berbagai industri seperti keuangan, ritel, kesehatan, dan pendidikan. Sebenarnya, mempertahankan kehadiran di internet saat ini hampir menjadi keharusan untuk menjalankan organisasi yang menguntungkan, yang mengakibatkan peningkatan terus-menerus dalam jumlah *website* saat ini. Namun, metode yang semakin canggih justru digunakan untuk menyerang dan menipu pengguna menjadi mungkin oleh peningkatan teknis yang ada[1].

Ribuan situs baru dibuat setiap hari yang dimana mengumpulkan data pengguna melalui fungsi login. Banyaknya *website* membuat sulit untuk menentukan situs mana yang aman[2]. Para peneliti telah mengusulkan banyak metode berbeda untuk mengenali *Uniform Resource Locators* (*URL*) phishing. Metode yang paling mudah untuk memblokir URL phishing didasarkan pada *blacklist* atau *whitelist URL*. Efektivitas deteksi ini sangat bergantung pada daftar *URL*. Sistem akan memblokir akses ke *URL* atau memutuskan untuk mengizinkan akses ke *URL* berdasarkan apakah *URL* tersebut sudah termasuk dalam daftar yang ada. Metode ini memerlukan daftar *URL* yang panjang, yang dapat dibuat secara manual atau otomatis, dan perlu sering diperbarui. Jika tidak, *URL phishing* yang baru dibuat akan mampu untuk melewati perlindungan yang ada[3].

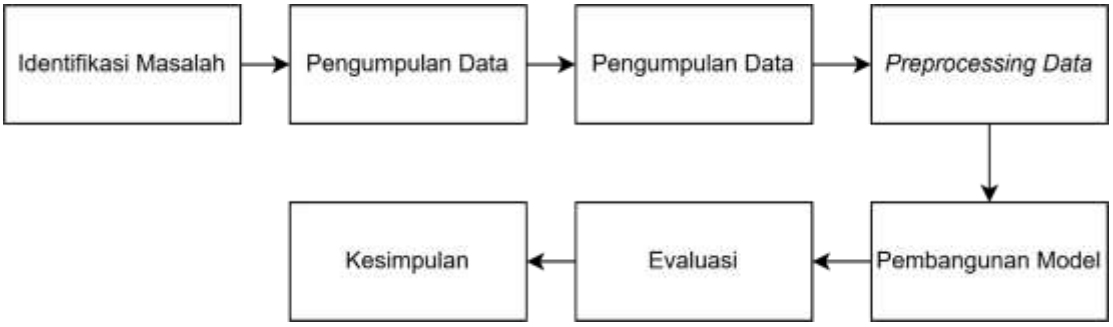
Metode konvensional seperti *blacklist URL* sering kali tidak cukup efektif karena tidak mampu mendeteksi URL baru yang belum terdaftar. Oleh karena itu, diperlukan solusi yang mampu melakukan deteksi secara cepat dan akurat terhadap *URL* berbahaya meskipun sebelumnya belum dikenal. Penelitian ini menawarkan solusi berbasis pembelajaran mesin dengan menggunakan algoritma Random Forest untuk mengklasifikasikan *URL phishing* berdasarkan 10

p-ISSN: 2986-3929
e-ISSN: 3032-1948

fitur struktural. Dengan hanya menggunakan informasi dari struktur URL itu sendiri tanpa perlu melakukan *crawling* atau akses eksternal, model ini diharapkan dapat bekerja secara *real-time* dan efisien, menjadikannya alternatif yang praktis untuk sistem keamanan web yang lebih adaptif.

2. Metode Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang dirancang secara sistematis yaitu meliputi identifikasi masalah, proses pengumpulan data, pemilihan fitur struktural, *preprocessing data*, pembangunan model menggunakan algoritma yang digunakan, evaluasi terhadap performa model, dan langkah terakhir adalah pengambilan kesimpulan dari model yang sudah dibuat. Untuk gambaran dari tahapan penelitian dapat diperhatikan pada Gambar 1 dibawah.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari Kaggle:
https://www.kaggle.com/eswarchandt/phishing-website-detector

Tabel 1 di bawah ini memberikan penjelasan yang lebih rinci mengenai hal ini.

Tabel 1. Sampel Data URL Berdasarkan Fitur Struktural

ID	Having IP Addreses	URL Length	Shortining Service	Having at Symbol	Result
1	-1	1	-1	-1	-1
2	1	0	1	1	1
3	-1	-1	-1	-1	-1
4	1	1	1	1	1
5	-1	-1	-1	-1	-1

Tabel di atas menampilkan lima entri awal dari data yang digunakan, terdiri atas empat fitur struktural dan satu label target (*Result*) yang menunjukkan klasifikasi URL sebagai *phishing* dengan indikasi nilai (-1), untuk aman dengan indikasi nilai (1), dan untuk netral dengan indikasi nilai (0).

2.2. Seleksi Fitur Struktural

Di dalam pemilihan *URL* biasanya melibatkan ekstraksi fitur atau penyisipan karakter. Ekstraksi fitur mengidentifikasi atribut URL penting, seperti *domain*, panjang *URL*, dan jumlah karakter, untuk dimasukkan ke dalam algoritma klasifikasi[4].

Oleh karena itu, seleksi fitur struktural dilakukan untuk menentukan fitur-fitur yang relevan dalam klasifikasi *URL* berbahaya tanpa bergantung pada konten halaman web atau informasi eksternal lainnya. Fitur struktural merupakan karakteristik yang dapat diekstrak langsung dari struktur *UR*L maupun elemen teknis situs tanpa perlu memuat konten halaman. Contoh fitur ini meliputi panjang *URL*, penggunaan simbol "@", keberadaan *redirect* tidak normal, serta apakah *URL* menggunakan protokol *HTTPS*. Dalam penelitian ini, dipilih sebanyak 10 fitur struktural dari total 30 fitur yang tersedia dalam *dataset*, yaitu: *UsingIP*, *LongURL*, *ShortURL*, *Symbol@*, *Redirecting//*, *HTTPS*, *AnchorURL*, *DNSRecording*, *AgeofDomain*, dan *AbnormalURL*. Pemilihan dilakukan dengan mempertimbangkan efisiensi, agar model yang dibangun tetap ringan namun efektif untuk mendeteksi *URL phishing* secara *real-time*.

2.3. *Preprocessing Data*

Prepocessing Data dilakukan untuk mempersiapkan data agar siap digunakan dalam proses pelatihan model. Tahapan ini mencakup pemilihan fitur struktural yang relevan, penghapusan fitur yang tidak diperlukan, serta pemeriksaan terhadap nilai yang hilang atau tidak konsisten. Seluruh fitur yang digunakan dalam penelitian ini telah berbentuk kategorikal numerik dengan nilai -1, 0, dan 1, sehingga tidak diperlukan proses *encoding* tambahan. Selain itu, dataset dibagi menjadi dua bagian, yaitu data latih dan data uji, dengan perbandingan 80:20 menggunakan teknik *stratified split* untuk menjaga proporsi kelas *phishing* dan *non-phishing* seimbang. Langkah ini bertujuan memastikan bahwa model yang dibangun dapat belajar secara optimal dan diuji dengan data yang merepresentasikan distribusi kelas secara adil.

2.4. Pembuatan Model

a. Algoritma Random Forest

Random Forest adalah himpunan algoritma pembelajaran terbimbing yang digunakan untuk klasifikasi dan regresi yang digunakan dalam pemodelan prediktif dan teknik pembelajaran *machine learning*. Random Forest menarik perhatian akademisi karena kecepatan dan keakuratannya dalam kategorisasi. Random Forest membagi kumpulan data menjadi dua bagian yaitu pelatihan (*training*) dan pengujian (*testing*). Random Forest digunakan untuk mengklasifikasikan *URL* itu terindikasi phishing atau tidak berdasarkan fitur – fitur struktural yang diekstraksi langsung dari bentuk *URL*. Random Forest dipilih karena mampu menangani data kategorikal dengan baik, relatif tahan terhadap *overfitting*, dan dapat memberikan informasi mengenai kontribusi setiap fitur terhadap hasil prediksi[5].

b. Parameter Model Random Forest

Model Random Forest dalam penelitian ini dibangun secara bertahap menggunakan pendekatan *warm_start*, di mana pelatihan dimulai dengan satu *decision tree* dan kemudian ditingkatkan secara progresif hingga mencapai total 50 *decision tree* (*n_estimators=50*). Pendekatan ini akan memantau kinerja model pada setiap tahap penambahan *tree* untuk menilai stabilitas akurasi. Parameter *max_depth=10* digunakan untuk membatasi kedalaman pohon dan mencegah *overfitting*. Pemisahan antar *node* menggunakan fungsi *Gini Index* (*criterion='gini'*), dan nilai acak diatur dengan *random_state=42* untuk memastikan reprodusibilitas hasil. Parameter lain seperti *min_samples_split* dan *min_samples_leaf* menggunakan nilai *default* karena telah sesuai dengan karakteristik data dan tidak menunjukkan kebutuhan untuk disesuaikan.

c. Voting Mayoritas dalam Random Forest

Random Forest menggunakan pendekatan voting mayoritas dalam proses klasifikasi. Setiap pohon keputusan yang membentuk model akan memberikan prediksi terhadap kelas dari suatu data yang dimasukkan, lalu hasil akhir ditentukan berdasarkan kelas yang paling banyak dipilih oleh seluruh pohon. Pendekatan ini bertujuan untuk mengurangi bias dari masing-masing pohon dan meningkatkan stabilitas serta akurasi model secara keseluruhan.

2.5. Evaluasi Model

Setelah model Random Forest berhasil dibangun menggunakan *training data*, langkah selanjutnya adalah melakukan evaluasi terhadap performa model dengan menggunakan *testing data*. Evaluasi dilakukan untuk mengukur seberapa baik model dalam mengklasifikasikan URL sebagai *phishing* atau tidak berdasarkan fitur-fitur struktural yang telah ditentukan. Beberapa metrik evaluasi yang digunakan dalam penelitian ini meliputi:

Dengan menggunakan keempat metrik ini, performa model dapat dinilai secara menyeluruh, baik dari segi ketepatan klasifikasi maupun kemampuannya dalam mendeteksi *URL phishing* secara akurat. Evaluasi dilakukan terhadap data uji yang telah dipisahkan sebelumnya dengan rasio 80:20 menggunakan teknik *stratified split*.

3. Hasil dan Diskusi

Model Random Forest yang dibangun dalam penelitian ini telah melalui proses pelatihan dan pengujian menggunakan data yang telah diproses dari *dataset URL phishing*. Evaluasi dilakukan untuk menilai sejauh mana model mampu mengklasifikasikan *URL* berbahaya dan tidak berbahaya secara akurat berdasarkan sepuluh fitur struktural yang digunakan. Hasil yang diperoleh menunjukkan bahwa model memiliki performa klasifikasi yang tinggi, ditunjukkan melalui nilai akurasi dan metrik evaluasi lainnya yang secara konsisten berada pada tingkat yang baik.

3.1. Hasil Evaluasi Model

a. Hasil Evaluasi Algoritma Random Forest

Berikut merupakan hasil evaluasi menggunakan *classification report*, model Random Forest menunjukkan performa yang sangat baik dalam mengklasifikasikan *URL phishing* dan *legitimate*. Untuk kelas *phishing* (-1), *precision* mencapai 97% dan *recall* sebesar 94%, yang berarti sebagian besar URL phishing berhasil dikenali dengan tingkat kesalahan yang rendah. Sementara itu, untuk kelas *legitimate* (1), *precision* tercatat sebesar 95% dan *recall* sebesar 98%, menunjukkan bahwa model mampu mendeteksi *URL* aman secara konsisten. Nilai *f1-score* untuk kedua kelas berada di atas 95%, yang mengindikasikan keseimbangan yang baik antara *precision* dan *recall*. Akurasi keseluruhan model terhadap data uji mencapai 96%. Untuk penjelasan lebih lengkap dapat dilihat pada gambar di bawah ini.

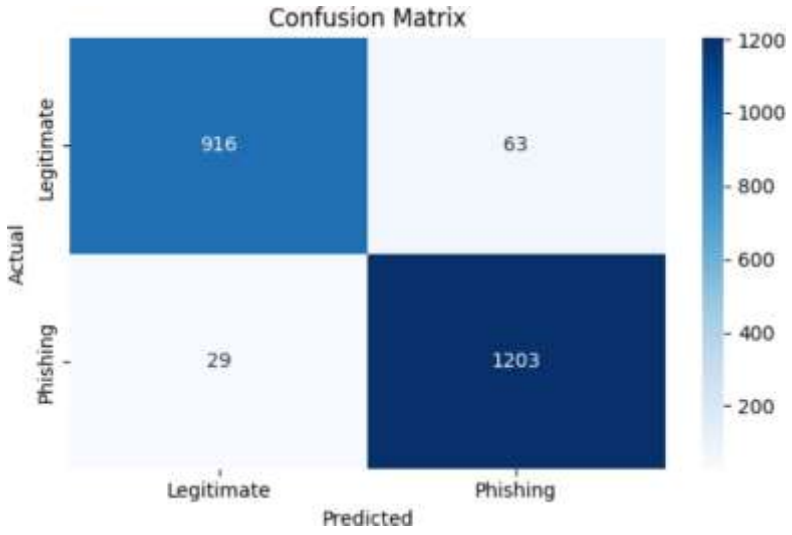
Classification Report:

	precision	recall	f1-score	support
-1	0.97	0.94	0.95	979
1	0.95	0.98	0.96	1232
accuracy			0.96	2211
macro avg	0.96	0.96	0.96	2211
weighted avg	0.96	0.96	0.96	2211

Gambar 2. Metrics Score Random Forest

Gambar di atas merupakan gambaran hasil penilaian *metrics score* dari algoritma Random Forest. Selanjutnya akan diperlihatkan untuk *confusion matrix* yang direpresentasikan dalam bentuk diagram matriks. Berikut merupakan visualisasi dari diagram matriks untuk *confusion matrix*.

p-ISSN: 2986-3929
e-ISSN: 3032-1948



Gambar 3. Diagram Matriks *Confusion Matrix*

Dari gambar 3 diatas, diketahui bahwa sebanyak 916 data *legitimate* berhasil diprediksi dengan benar (*True Negative*), sementara 1203 data *phishing* juga diklasifikasikan dengan tepat (*True Positive*). Namun, masih terdapat 63 data *legitimate* yang salah diklasifikasikan sebagai *phishing* (*False Positive*) dan 29 data *phishing* yang salah dikenali sebagai *legitimate* (*False Negative*). Dengan total data uji sebanyak *2211*, model berhasil melakukan prediksi yang benar pada 2119 data, menghasilkan tingkat akurasi sekitar 95,84%. *Confusion matrix* ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik, dengan tingkat kesalahan yang relatif rendah, sehingga cukup efektif dalam membedakan antara *URL phishing* dan *legitimate* berdasarkan fitur struktural yang digunakan.

b. Hasil Akurasi *testing* dan *testing*

Dalam penelitian ini, model Random Forest dievaluasi berdasarkan akurasi pada data pelatihan (*training*) dan data pengujian (*testing*). Hasil evaluasi menunjukkan bahwa akurasi pada data *training* mencapai 92,41%, sedangkan akurasi pada data *testing* sebesar 92,72%. Berikut untuk penjelasan lebih lengkap dapat dilihat pada tabel 2 di bawah ini.

Tabel 2. Perbandingan akurasi *training* dan *testing*

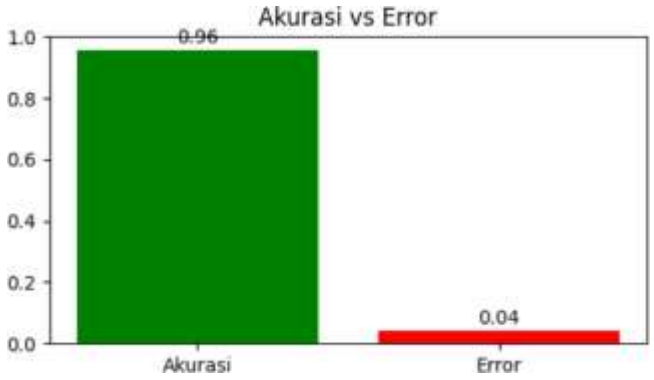
Jenis	Akurasi	Persentase
Testing	0.9272	92,72%
Training	0.9241	92,41%

Tabel di atas memperlihatkan perbandingan antara data *training* dan *testing* yang menunjukkan bahwa model tidak mengalami *overfitting* atau *underfitting*.

c. Hasil Perbandingan Akurasi dan *Error*

Dalam penelitian ini, model mampu untuk menampilkan nilai akurasi dan *error* dari model Random Forest pada data pengujian. Akurasi menunjukkan proporsi prediksi yang benar, sementara *error* merupakan kebalikannya, yaitu proporsi prediksi yang salah. Berikut untuk penjelasan lebih lengkap dapat dilihat pada gambar 4 di bawah ini.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

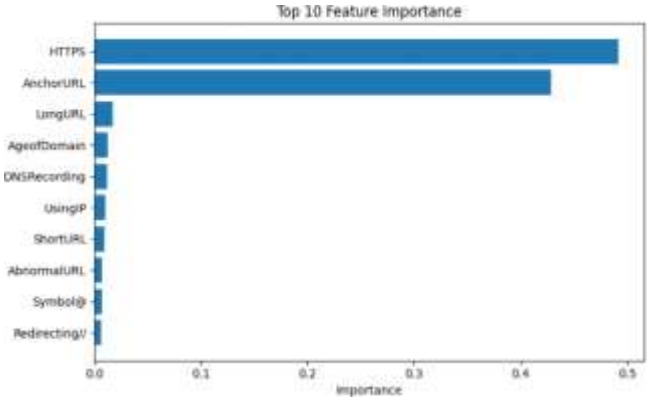


Gambar 4. Diagram batang perbandingan akurasi dan *error*

Gambar di atas memperlihatkan akurasi model mencapai 96%, sedangkan *error* hanya 4%, yang ditunjukkan melalui perbandingan diagram batang yang kontras antara keduanya. Visualisasi ini mempertegas bahwa model memiliki performa yang sangat baik dalam mendeteksi *URL phishing*, dengan tingkat kesalahan yang sangat rendah.

3.2. Fitur Paling Berpengaruh

Berdasarkan hasil pelatihan model Random Forest menggunakan 10 fitur struktural, diketahui bahwa beberapa fitur memiliki kontribusi yang lebih signifikan dalam mempengaruhi hasil klasifikasi. Berikut untuk penjelasan lebih lengkap dapat dilihat pada gambar 5 di bawah ini.



Gambar 5. Grafik *feature importance*

Berdasarkan gambar di atas, fitur yang paling berpengaruh dalam klasifikasi *URL phishing* adalah *HTTPS* dan *AnchorURL*. Kedua fitur ini menunjukkan tingkat penting yang jauh lebih tinggi dibandingkan fitur lainnya. Fitur HTTPS mencerminkan apakah URL menggunakan protokol aman, yang secara signifikan berhubungan dengan keamanan situs. *AnchorURL* mengindikasikan apakah tautan di dalam halaman mengarah ke sumber yang mencurigakan atau tidak, yang sering menjadi ciri khas halaman *phishing*. Sementara itu, fitur lain seperti *LongURL*, *AgeofDomain*, dan *DNSRecording* masih berkontribusi terhadap prediksi, namun dalam skala yang jauh lebih kecil. Fitur seperti *UsingIP*, *ShortURL*, *AbnormalURL*, *Symbol@*, dan *Redirecting//* berada pada level yang rendah dalam model, namun tetap diperlukan karena kontribusinya tetap membantu untuk meningkatkan akurasi model secara keseluruhan.

4. Kesimpulan

Model Random Forest yang digunakan dalam penelitian ini berhasil menunjukkan performa yang sangat baik dalam mendeteksi *URL phishing* dengan menggunakan 10 fitur struktural yang dipilih secara selektif. Model dibangun dengan 50 pohon keputusan (*n_estimators=50*) dan kedalaman

maksimum 10 (*max_depth=10*) untuk menjaga keseimbangan antara akurasi dan kompleksitas, serta menghindari *overfitting*. Berdasarkan hasil evaluasi, model mencapai akurasi tinggi sebesar 96% pada data pengujian, dengan nilai *precision* dan *recal*l yang seimbang pada masing-masing kelas. Fitur paling berpengaruh terhadap keputusan model adalah kehadiran protokol *HTTPS* dan karakteristik *anchor URL*, yang menunjukkan bahwa indikator keamanan dan struktur tautan menjadi faktor penting dalam identifikasi *URL phishing*. Visualisasi *confusion matrix* dan diagram batang akurasi vs *error* semakin memperkuat bahwa model memiliki kemampuan klasifikasi yang stabil untuk diterapkan dalam sistem klasifikasi *phishing* secara *real-time*.

Daftar Pustaka

[1] L. Eze, U. B. Chaudhry, and H. Jahankhani, “Quantum-Enhanced Machine Learning for Cybersecurity: Evaluating Malicious URL Detection,” *Electronics (Switzerland)*, vol. 14, no. 9, May 2025, doi: 10.3390/electronics14091827.

[2] S. Abad, H. Gholamy, and M. Aslani, “Classification of Malicious URLs Using Machine Learning,” *Sensors*, vol. 23, no. 18, Sep. 2023, doi: 10.3390/s23187760.

[3] S. R. Abdul Samad *et al.*, “Analysis of the Performance Impact of Fine-Tuned Machine Learning Model for Phishing URL Detection,” *Electronics (Switzerland)*, vol. 12, no. 7, Apr. 2023, doi: 10.3390/electronics12071642.

[4] M. Y. Su and K. L. Su, “BERT-Based Approaches to Identifying Malicious URLs,” *Sensors (Basel)*, vol. 23, no. 20, Oct. 2023, doi: 10.3390/s23208499.

[5] S. Alnemari and M. Alshammari, “Detecting Phishing Domains Using Machine Learning,” *Applied Sciences (Switzerland)*, vol. 13, no. 8, Apr. 2023, doi: 10.3390/app13084649.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Halaman ini sengaja dibiarkan kosong