JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 14 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v14i3.9680

KLASIFIKASI LINK PHISHING PADA QR CODE MENGGUNAKAN ARSITEKTUR MULTI-TIER BERBASIS ALGORITMA RANDOM FOREST

Daffa Alde Rayhan Nurdiyanto Putra[1*], Supatman[2]

[1,2]Universitas Mercu Buana Yogyakarta; Jl. Ring Road Utara, Ngropoh, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281; Telp: 0274-2801918

Keywords:
Phishing;
Random Forest;
Machine Learning;
Klasifikasi.

Corespondent Email:
221110009@student.mercubuana-yogya.ac.id

Abstrak. Transformasi digital mengubah interaksi siber namun memicu ancaman rekayasa sosial baru seperti quishing atau QR Phishing. Metode ini memanfaatkan kemudahan kode QR untuk menyamarkan tautan berbahaya dari sistem keamanan otomatis. Keterbatasan metode blacklist dan model klasifikasi selapis terhadap manipulasi visual mendesak adanya solusi pertahanan yang lebih fleksibel dan komprehensif. Penelitian ini menerapkan sistem deteksi multi-tier berbasis algoritma Random Forest yang dilengkapi kemampuan pemindaian kode QR untuk memperluas jangkauan deteksi terhadap ancaman siber. Metodologi yang digunakan mencakup ekstraksi fitur leksikal URL, penggunaan Trusted Link Dataset yang terdiri dari 12 juta entri, serta pengolahan citra QR untuk mengekstrak tautan tersembunyi. Hasil pengujian menunjukkan bahwa model Random Forest yang diterapkan dalam sistem ini mampu mencapai tingkat akurasi sebesar 86%. Integrasi mekanisme deteksi berlapis ini terbukti efektif dalam memitigasi berbagai vektor serangan siber modern yang tidak terdeteksi oleh sistem statis, sekaligus memberikan kontribusi nyata bagi pengembangan sistem keamanan siber yang lebih kuat melalui pendekatan klasifikasi cerdas.

Abstract. *Digital transformation is reshaping cyber interactions but has also given rise to new social engineering threats such as quishing, or QR phishing. This method exploits the convenience of QR codes to disguise malicious links from automated security systems. The limitations of blacklisting methods and single-layer classification models against visual manipulation underscore the urgent need for more flexible and comprehensive defense solutions. This research implements a multi-tier detection system based on the Random Forest algorithm, equipped with QR code scanning capabilities to expand the detection scope against cyber threats. The methodology used includes URL lexical feature extraction, the use of the Trusted Link Dataset consisting of 12 million entries, and QR image processing to extract hidden links. Test results show that the Random Forest model implemented in this system achieves an accuracy rate of 86%. The integration of this multi-layered detection mechanism has proven effective in mitigating various modern cyberattack vectors undetected by static systems, while making a tangible contribution to the development of stronger cybersecurity systems through an intelligent classification approach.*

1. PENDAHULUAN

Transformasi digital yang masif telah mengubah paradigma interaksi manusia di ruang siber secara fundamental, namun di sisi lain turut menghadirkan ancaman keamanan yang semakin kompleks dan dinamis sehingga memerlukan perhatian serius [1]. Salah satu ancaman paling dominan yang terus berevolusi adalah *phishing*, yakni teknik rekayasa sosial berbasis manipulasi URL yang bertujuan mencuri kredensial maupun data sensitif lainnya [2]. Serangan ini dilakukan dengan menduplikasi identitas otoritas tepercaya, seperti institusi perbankan atau platform *e-commerce*, guna mengarahkan korban ke infrastruktur palsu yang menyerupai entitas legal [3]. Melalui penetrasi tersebut, aktor ancaman dapat memantau sekaligus memanipulasi lalu lintas komunikasi dalam jaringan yang berpotensi menimbulkan kerugian material serta kebocoran data privasi dalam skala besar [4].

Seiring dengan evolusi tersebut, muncul sebuah varian ancaman baru yang dikenal sebagai *quishing*, yakni bentuk kejahatan siber yang menggunakan kode QR berbahaya untuk menipu pengguna dengan memanfaatkan kepercayaan serta kemudahan penggunaan teknologi tersebut dalam kehidupan digital sehari-hari [5]. dalam studi lapangan terhadap 173 partisipan bahwa sebanyak 67% subjek dengan mudah memberikan kredensial Google atau Facebook mereka, di mana faktor kemudahan menjadi alasan utama kerentanan tersebut [6]. Yang lebih mengkhawatirkan, dalam simulasi *phishing* skala besar dengan lebih dari 71.000 email menemukan bahwa email *quishing* memiliki tingkat keberhasilan yang setara dengan *phishing* tradisional dalam mengarahkan pengguna ke halaman web berbahaya, namun jauh lebih sulit dideteksi oleh sistem keamanan otomatis [7]. Kejahatan ini secara strategis memanfaatkan kelemahan psikologis seperti bias kepercayaan lingkungan dan kebiasaan memindai kode QR secara otomatis, serta mengeksploitasi teknik teknis seperti perumitan URL dan pembajakan sesi [8].

Untuk memitigasi risiko tersebut, penguatan aspek teknis dan perilaku pengguna menjadi faktor fundamental. Langkah preventif konvensional seperti pembaruan sistem berkala, penggunaan *firewall*, dan implementasi *Intrusion Detection System* (IDS) terbukti mampu meningkatkan pertahanan awal [9], [10]. Namun, untuk menghadapi pola ancaman yang responsif, pemanfaatan kecerdasan buatan melalui algoritma *machine learning* menjadi solusi strategis dalam membangun sistem deteksi yang otomatis dan akurat.

Di antara berbagai algoritma klasifikasi, *Random Forest* secara konsisten menunjukkan performa superior dalam mengidentifikasi situs web berbahaya. Sebagai algoritma *ensemble learning* yang bekerja melalui mekanisme *majority voting* dari sekumpulan pohon keputusan (*decision trees*), *Random Forest* menawarkan stabilitas tinggi dan meminimalkan risiko *overfitting*. Penelitian terbaru menunjukkan algoritma ini mampu mencapai akurasi hingga 97,2% [11], mengungguli metode lain seperti *Decision Tree dan K-Nearest Neighbors* [12]. Integrasi metode *stacking* bahkan dapat meningkatkan akurasinya hingga 96,4% pada *dataset binary class* [13], menjadikannya model yang paling andal untuk menangani fitur kompleks pada URL berbahaya.

Meskipun *Random Forest* terbukti efektif, mayoritas studi terdahulu masih terpaku pada mekanisme deteksi satu lapis (*single-tier*) yang hanya mengevaluasi fitur leksikal URL secara statis. Keterbatasan ini menciptakan celah keamanan terhadap teknik pengelabuan visual seperti *quishing*. Dalam modus ini, tautan berbahaya disamarkan dalam kode QR yang tidak dapat dijangkau oleh algoritma deteksi teks konvensional [14].

Oleh karena itu, terdapat kebutuhan mendesak untuk mendiversifikasi metode deteksi. Integrasi mekanisme *multi-tier* yang mencakup dekripsi kode QR dan verifikasi fitur struktural secara bertahap menjadi krusial untuk mengatasi kompleksitas serangan *quishing* dan memastikan ketahanan sistem terhadap manipulasi URL yang melampaui batas deteksi statis [15].

2. TINJAUAN PUSTAKA

2.1. Phishing

Phishing didefinisikan sebagai salah satu bentuk kejahatan siber yang memanfaatkan teknik manipulasi psikologis untuk mengelabui target agar memberikan data sensitif, mencakup informasi identitas pribadi, kredensial akun, hingga data finansial yang bersifat rahasia [16]. Aktivitas ini terbukti dapat dilakukan dengan tingkat kemudahan yang tinggi melalui infrastruktur internet, sehingga menjadi katalis bagi munculnya berbagai varian kejahatan daring lainnya.

Seiring dengan perkembangan teknologi informasi, metode pendistribusian serangan *phishing* semakin beragam dan masif, mulai dari pemanfaatan surat elektronik hingga manipulasi antarmuka situs web yang menyerupai aslinya. Fenomena ini menimbulkan risiko kerugian materiil maupun non-materiil yang signifikan bagi pengguna. Oleh karena itu, pemahaman mendalam mengenai karakteristik serangan *phishing* menjadi krusial untuk membangun sistem pertahanan yang adaptif. Efektivitas deteksi terhadap ancaman ini diperlukan guna memitigasi dampak dari perluasan spektrum kejahatan siber di ruang digital.

2.2. Machine Learning

Machine Learning adalah suatu cabang ilmu komputer yang memungkinkan komputer untuk belajar dari data dan mengembangkan model yang dapat membuat prediksi atau keputusan berdasarkan pengalaman tersebut. Dengan menggunakan algoritma dan teknik yang kompleks, *Machine Learning* dapat membantu dalam berbagai bidang, seperti pendidikan, kesehatan, dan bisnis, untuk meningkatkan efisiensi dan akurasi dalam pengambilan keputusan [17].

2.3. Random Forest

Random Forest merupakan salah satu algoritma klasifikasi yang berbasis pada teknik *ensemble learning*. Algoritma ini bekerja dengan menggabungkan sejumlah model klasifikasi sederhana, seperti *Decision Tree*, untuk menghasilkan prediksi yang lebih stabil dan akurat [18]. Dengan memanfaatkan prinsip voting dari banyak pohon keputusan, *Random Forest* mampu mengurangi risiko kesalahan prediksi yang biasanya muncul pada model tunggal.

Keunggulan utama *Random Forest* terletak pada kemampuannya menangani data berukuran besar dan kompleks, serta mengurangi potensi *overfitting* yang sering terjadi pada algoritma klasifikasi lain. Selain itu, algoritma ini fleksibel dalam mengolah berbagai jenis fitur, baik numerik maupun kategoris, sehingga banyak digunakan dalam penelitian keamanan siber, termasuk deteksi *phishing*. Studi terdahulu menunjukkan bahwa *Random Forest* dapat mencapai tingkat akurasi tinggi dalam membedakan tautan *phishing* dan tautan sah, menjadikannya salah satu pendekatan yang relevan untuk pengembangan sistem klasifikasi berbasis URL.

2.4. Kode QR

Kode QR merupakan citra dua dimensi yang dirancang untuk menyimpan dan menyampaikan informasi secara ringkas. Format ini mampu menampung berbagai jenis data, termasuk teks numerik, alfanumerik, maupun representasi biner [19].

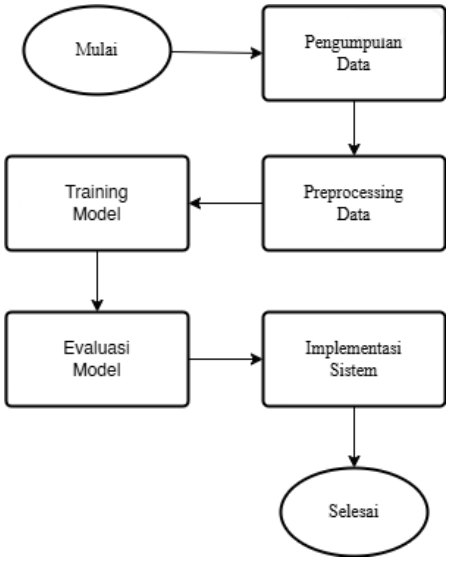
2.5. Penelitian Terkait

Beberapa penelitian terdahulu telah mengkaji penerapan algoritma *Random Forest* dalam klasifikasi *link phishing* berbasis web. Pendekatan berbasis analisis fitur leksikal pada URL dilaporkan mampu menghasilkan tingkat akurasi yang tinggi dalam membedakan situs legal dan situs ilegal. Namun demikian, sebagian besar penelitian tersebut masih berfokus pada klasifikasi teks secara langsung tanpa melibatkan mekanisme validasi berlapis yang mampu menangani evolusi taktik penyerangan terbaru, seperti penggunaan kode QR.

Selain itu, kajian mengenai pemanfaatan daftar situs yang sudah dikonfirmasi aman sebagai penyaring awal sebelum proses *machine learning* masih sangat terbatas. Kondisi ini menunjukkan adanya kebutuhan untuk mengembangkan sistem deteksi multifaktor yang tidak hanya mengandalkan satu jalur klasifikasi, tetapi juga mengintegrasikan pemindaian kode QR dan validasi terhadap sumber tepercaya.

3. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan utama yang dirancang sedemikian rupa agar hasil deteksi memiliki tingkat akurasi yang optimal, sebagaimana ditunjukkan dalam diagram alur berikut.



Gambar 1. Alur Proses Penelitian

3.1. Pengumpulan Data

Tahap akuisisi data dilakukan dengan menghimpun dua *dataset* sekunder yang bersumber dari repositori *Kaggle* untuk mendukung mekanisme deteksi pada sistem. *Dataset* pertama terdiri dari sekitar 17 juta data domain global yang difungsikan sebagai *Trusted Link Dataset* untuk memvalidasi situs-situs yang telah terverifikasi aman secara instan. *Dataset* kedua digunakan untuk kebutuhan pelatihan dan pengujian model *machine learning*, yang mencakup 641.119 sampel URL dengan label kategori *benign*, *phishing*, *defacement* dan *malware*. Penggunaan kedua *dataset* berskala besar ini bertujuan untuk membangun sistem yang komprehensif, di mana *Trusted Link Dataset* berperan dalam mengoptimalkan kecepatan respons, sementara *dataset* pelatihan memberikan keragaman data yang luas bagi algoritma *Random Forest* untuk mendeteksi pola *phishing* secara akurat.

3.2. Preprocessing Data

Tahap pra-pemrosesan diawali dengan pembersihan *Trusted Link Dataset* melalui penghapusan kolom non-esensial dan entitas kosong, yang mereduksi data dari 17 juta menjadi 12 juta domain valid. Pada *dataset* pelatihan, dilakukan penyederhanaan label dari empat kategori menjadi klasifikasi biner, yaitu *safe* dan *suspicious*, untuk mengoptimalkan fokus identifikasi model. Selanjutnya, dilakukan ekstraksi fitur leksikal untuk mentransformasi URL menjadi data numerik, yang mencakup perhitungan panjang URL, panjang domain, jumlah digit, serta keberadaan simbol khusus seperti titik, tanda hubung, garis miring, dan simbol '@'. Proses ini memungkinkan algoritma untuk melakukan analisis komparatif berdasarkan karakteristik struktural dari setiap tautan yang diproses.

Table 1.Sampel Dataset Training

Url	Tipe
br-icloud.com.br	*suspicious*
mp3raid.com/music/krizz_kaliko.html	*safe*
bopsecrets.org/rexroth/cr/1.htm	*safe*
www.garage-pirenne.be/index.php?option=com_content&view=article&id=70&vsig70_0=15	*suspicious*

3.3. Training Model

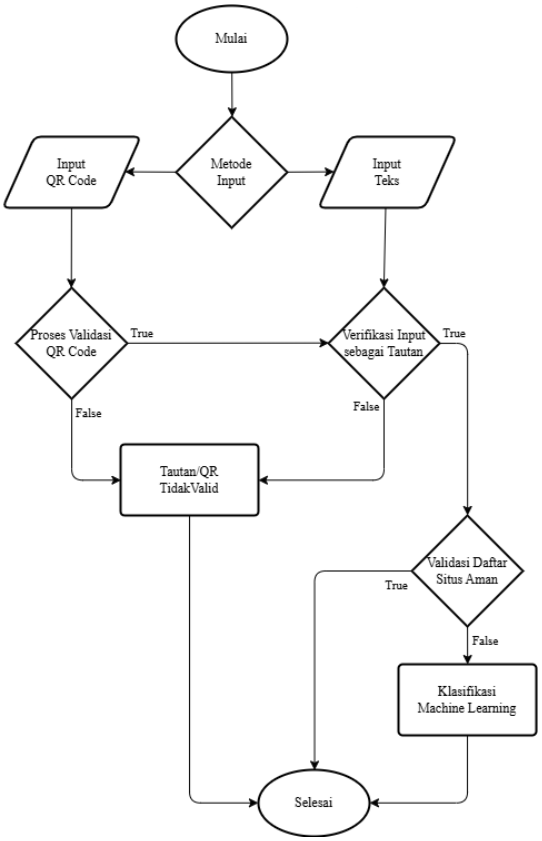
Pelatihan ini dilakukan secara eksklusif menggunakan *dataset* pelatihan (*training dataset*) yang telah diproses sebelumnya, bukan menggunakan *Trusted Link Dataset*. *Dataset* tersebut dibagi dengan rasio 80% untuk data latih dan 20% untuk data uji guna meminimalkan risiko *overfitting*. Selama proses pelatihan, algoritma melakukan *random sampling* pada fitur dan data untuk membangun model yang variatif. Penentuan hasil akhir klasifikasi didasarkan pada mekanisme *majority voting* dari seluruh pohon yang terbentuk, sehingga model memiliki tingkat stabilitas dan ketahanan yang lebih tinggi dibandingkan penggunaan pohon keputusan tunggal.

3.4. Evaluasi Model

Proses evaluasi ini dilakukan menggunakan dua metode standar, yaitu *classification report* dan *confusion matrix*. Penggunaan *classification report* bertujuan untuk mendapatkan metrik penilaian yang komprehensif, di mana setiap parameter dihitung berdasarkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Sementara *confusion matrix* digunakan untuk memvisualisasikan performa model secara detail dengan membandingkan nilai prediksi terhadap nilai aktual.

3.5. Implementasi Sistem

Implementasi sistem deteksi *phishing* dalam penelitian ini dirancang menggunakan pendekatan *multi-layered filtering* untuk mengoptimalkan waktu pemrosesan dan akurasi deteksi. Alur kerja sistem secara sistematis digambarkan pada Gambar 3. Secara mendalam, tahapan eksekusi sistem dijelaskan sebagai berikut:



Gambar 2. Alur Eksekusi Sistem

Tahap awal dalam Implementasi Sistem ini berfokus pada mekanisme akuisisi data yang diintegrasikan melalui dua modalitas input utama, yakni pemindaian QR Code dan entri teks secara manual. Pada jalur pertama, sistem melakukan pemrosesan citra untuk memvalidasi integritas struktur kode QR, guna memastikan bahwa data yang dipindai memenuhi standar sintaksis yang dapat terbaca. Secara paralel, pada jalur entri teks, dilakukan prosedur klasifikasi awal untuk mengidentifikasi apakah entitas data merupakan tautan atau teks non-tautan. Tahapan ini berfungsi sebagai filter primer dalam memastikan bahwa input yang masuk ke dalam sistem memiliki format yang sesuai sebelum dilakukan analisis pada level yang lebih kompleks.

Pada tahap pemrosesan inti, tautan yang telah teridentifikasi akan disaring secara pasti dengan mencocokkannya terhadap data *Trusted Link Dataset* yang telah disiapkan sebelumnya. *Dataset* ini bersumber dari *BigPicture* , yang menyediakan repositori data entitas global terverifikasi mencakup lebih dari 17 juta domain perusahaan yang telah diperbarui pada kuartal kedua tahun 2024. Proses pencocokan menggunakan referensi data legal dan korporasi yang kredibel ini krusial untuk memisahkan entitas yang telah terverifikasi kredibilitasnya dari entitas yang belum dikenal, sehingga mempercepat waktu eksekusi sistem pada data yang sudah dianggap aman oleh basis data internal.

Pada tahap akhir, untuk data yang memerlukan evaluasi lebih mendalam, sistem menjalankan prosedur klasifikasi berbasis *Machine Learning* Tahapan ini dirancang untuk mendeteksi pola-pola anomali dan potensi ancaman secara dinamis yang tidak mampu dikenali oleh filter statis pada tahap sebelumnya. Integrasi antara pengecekan *Trusted Link Dataset* dan analisis prediktif cerdas ini membentuk satu kesatuan sistem validasi berlapis. Setelah seluruh rangkaian prosedur, mulai dari akuisisi, verifikasi protokol, hingga analisis algoritma cerdas terpenuhi, sistem akan mencapai titik terminasi atau status selesai, yang menandakan bahwa data input telah melalui seluruh filter keamanan dan validitas yang dipersyaratkan.

4. HASIL DAN PEMBAHASAN

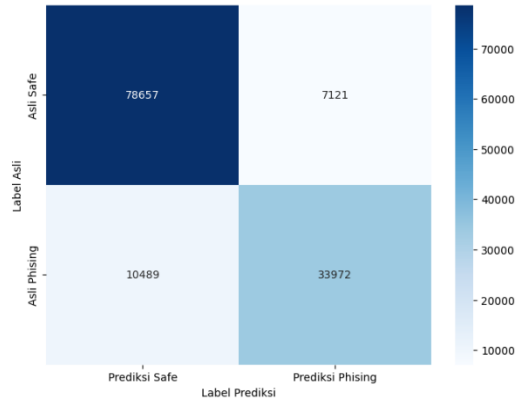
4.1 Analisis Performa Model Klasifikasi

Tahap evaluasi dilakukan untuk mengukur performa model dalam mengklasifikasikan tautan *phishing* dan *safe* berdasarkan data uji yang telah ditentukan. Analisis performa ini disajikan melalui metrik *Classification Report* yang mencakup nilai *precision, recall,* dan *f1-score*, serta *Confusion Matrix* untuk memberikan gambaran visual mengenai distribusi prediksi model terhadap label aktual data.

Table 2. Classification Report

Class	*Precission*	*Recall*	*F1-Score*
Safe	0.88	0.92	0.90
Suspicious	0.83	0.76	0.79
Accuracy			0.86

Pengujian model menggunakan 130.239 data uji menghasilkan akurasi sebesar 86%. Berdasarkan *classification report*, kelas Safe memperoleh *precision* 0,88 dan *recall* 0,92, sedangkan kelas *Suspicious* mencatat *precision* 0,83 dan *recall* 0,76. Performa ini menunjukkan model lebih unggul dalam mengenali situs aman dibandingkan situs *phishing*, dengan nilai F1-score rata-rata sebesar 0,85 yang mengonfirmasi stabilitas klasifikasi secara makro.



Gambar 3. Confusion Matrix

Analisis *confusion matrix* memperlihatkan bahwa model berhasil memprediksi 78.657 data Safe dan 33.972 data *Phishing* secara akurat. Namun, tercatat adanya 7.121 kasus *false positive* dan 10.489 *false negative*. Tingginya angka *false negative* tersebut menjadi landasan pentingnya fitur tambahan pada sistem web, yaitu mekanisme *Trusted Link Dataset* dan validasi status URL, guna memperkuat lapisan deteksi yang tidak terjangkau oleh model klasifikasi tunggal dalam skenario *real-time*.

4.2 Evaluasi Efektivitas Implementasi Sistem

Pengujian pada bagian ini bertujuan untuk menilai sejauh mana sistem deteksi *phishing* yang diimplementasikan mampu meningkatkan akurasi dibandingkan kondisi tanpa implementasi sistem. Evaluasi dilakukan dengan menggunakan sejumlah sampel data uji yang terdiri atas *trusted*,*safe*,*suspicous* dan *not a link*.

Table 3. Sampel Pengujian Tautan

Input Data	Status
google.com	*trusted*
daffalde.site	*safe*
vanillacafes.com/matcha-latte/	*suspicous*
wifipassword	*not a link*

Data tersebut diproses melalui dua skenario, yaitu dengan implementasi sistem yang telah diintegrasikan dan tanpa sistem (*baseline*). Hasil pengujian kemudian dibandingkan untuk melihat perbedaan tingkat akurasi serta efektivitas deteksi.

Table 4. Hasil Pengujian

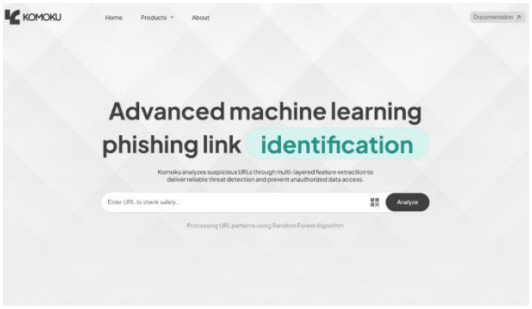
Tautan	*Baseline*	Implementasi Sistem
google.com	Gagal	Berhasil
daffalde.site	Berhasil	Berhasil
vanillacafes.com/matcha-latte/	Gagal	Berhasil
wifipassword	Berhasil	Berhasil

Pengujian menunjukkan bahwa pada skenario *Baseline* sebagian besar tautan berhasil diklasifikasikan dengan benar, seperti daffalde.site, dan wifipassword. Namun, terdapat kasus tertentu seperti google.com dan vanillacafes.com/matcha-latte/ yang gagal terdeteksi, sehingga menimbulkan kesalahan klasifikasi. Kondisi ini memperlihatkan bahwa pendekatan murni berbasis *machine learning* masih memiliki keterbatasan dalam menangani variasi data tertentu.

Sebaliknya, pada skenario Implementasi Sistem, seluruh tautan yang diuji berhasil diklasifikasikan dengan tepat, termasuk tautan yang sebelumnya gagal pada *baseline*. Hal ini menunjukkan bahwa integrasi sistem multifaktor memberikan peningkatan konsistensi dan keandalan dalam proses deteksi. Dengan demikian, hasil perbandingan ini menegaskan efektivitas sistem yang diusulkan dalam mengurangi kesalahan klasifikasi dan memperkuat performa deteksi *phishing* secara keseluruhan.

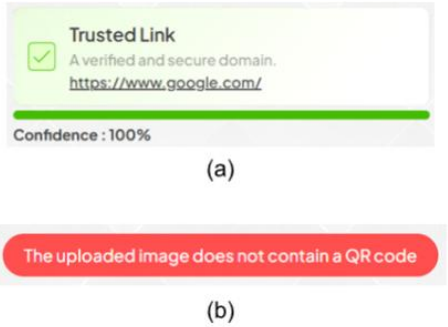
4.3 Pengujian Fungsionalitas

Implementasi arsitektur sistem ini mengintegrasikan *FastAPI* pada sisi *backend* untuk mengoptimalkan kinerja pemrosesan data asinkron serta eksekusi model *machine learning* secara efisien. Pada lapisan antarmuka, digunakan *framework Next.js* berbasis *TypeScript* guna menjamin keamanan tipe serta keandalan interaksi pengguna.



Gambar 4. Tampilan Utama Website

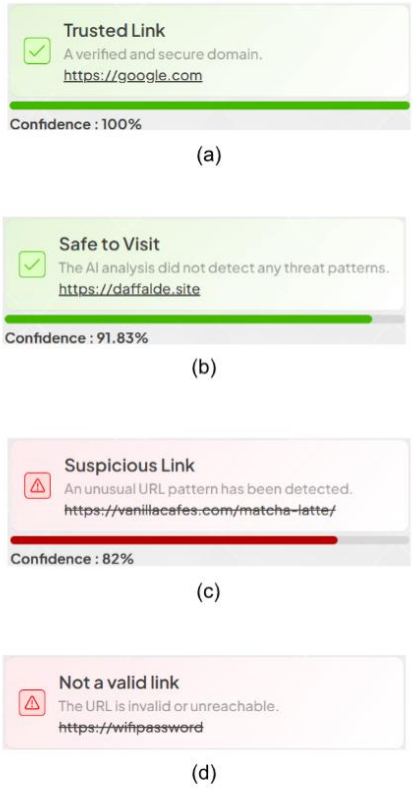
Pengujian awal dilakukan terhadap fitur pemindaian kode QR guna mengevaluasi fungsionalitas fitur pemindaian dilakukan melalui dua skenario pengujian utama guna memverifikasi keandalan sistem dalam mengenali input visual secara akurat. Pertama, pengujian menggunakan kode QR yang berisi tautan (*link*) valid untuk mengukur kemampuan sistem dalam melakukan dekripsi serta ekstraksi URL dengan tepat. Kedua, pengujian dilakukan menggunakan citra digital yang tidak mengandung kode QR sama sekali untuk memastikan efektivitas sistem dalam mengidentifikasi serta menolak input yang tidak relevan secara otomatis sebelum memasuki tahap pemrosesan data lebih lanjut.



Gambar 5. Hasil Pengujian Kode QR

Pada skenario pertama (a), sistem berhasil melakukan dekripsi terhadap kode QR yang mengandung tautan aktif dan secara presisi mengklasifikasikannya sebagai "*Trusted Link*" dengan tingkat kepercayaan (*confidence*) mencapai 100%. Sementara itu, hasil pengujian pada hasil (b) mengonfirmasi ketangguhan mekanisme validasi sistem, di mana sistem mampu secara otomatis mendeteksi serta menolak unggahan citra digital yang tidak memiliki komponen kode QR dengan menampilkan pesan peringatan yang informatif. Hal ini membuktikan bahwa sistem memiliki integritas fungsional yang baik dalam membedakan antara input yang valid untuk diproses lebih lanjut dan input yang tidak relevan, sehingga menjamin efisiensi alur kerja pada lapisan analisis keamanan berikutnya.

Tahap berikutnya bertujuan untuk memvalidasi performa sistem dalam menangani berbagai variasi input secara *real-time* melalui skenario *black box* yang menguji modul validasi format, filtrasi *Trusted Link Dataset*, dan mesin inferensi *Random Forest*. Pengujian menggunakan sampel pengujian dari Tabel 3 sebelumnya.



Gambar 6. Hasil Klasifikasi URL

Berdasarkan hasil pengujian yang dipaparkan pada Gambar 6, sistem menunjukkan performa yang konsisten dalam mengeksekusi seluruh skenario validasi dan klasifikasi. Pada hasil (a), modul *Trusted Link Dataset* berhasil mengidentifikasi input sebagai *Trusted* Link dengan tingkat kepercayaan mutlak sebesar 100%. Selanjutnya, (b) menunjukkan kemampuan mesin inferensi *Random Forest* dalam menghasilkan klasifikasi *Safe to Visit* dengan skor kepercayaan 91,83%. Pada pengujian terhadap indikasi ancaman yang ditunjukkan di hasil (c), sistem secara akurat menetapkan kategori *Suspicious Link* dengan nilai kepercayaan 82%. Terakhir, (d) mengonfirmasi efektivitas validasi format dengan mendeteksi input non-URL sebagai *Not a valid link*.

5. KESIMPULAN

Penelitian ini telah berhasil mengembangkan sistem klasifikasi link *phishing* dengan mengintegrasikan algoritma *Random Forest* dan mekanisme pertahanan berlapis. Berdasarkan hasil pengujian eksperimental, model klasifikasi mencapai tingkat akurasi sebesar 86% dengan nilai *F1-score* sebesar 0,85, yang menunjukkan stabilitas performa dalam membedakan antara URL aman dan mencurigakan. Secara fungsional, integrasi sistem pada lingkungan web menunjukkan tingkat keberhasilan tinggi, di mana modul validasi format dan *Trusted Link Dataset* mampu bekerja sinkron dengan mesin inferensi *Random Forest*.

Kelebihan utama dari sistem ini terletak pada arsitektur pertahanan bertingkat yang efisien, di mana fitur *Trusted Link Dataset* mempercepat respons deteksi pada domain populer dan validasi format menjaga reliabilitas proses ekstraksi fitur secara *real-time*. Namun, sistem masih memiliki kekurangan signifikan berupa angka *false negative* sebanyak 10.489 data pada tahap eksperimen, yang mengindikasikan keterbatasan fitur leksikal dalam mendeteksi teknik *phishing* yang lebih canggih. Selain itu, sistem saat ini masih bersifat anonim dan searah, sehingga hasil deteksi belum tersimpan secara sistematis untuk kebutuhan analisis jangka panjang maupun peningkatan performa model secara mandiri.

Untuk pengembangan selanjutnya, terdapat beberapa poin yang dapat diimplementasikan guna meningkatkan efektivitas sistem:

a. Menambah klasifikasi ancaman siber lain seperti *Email* dan *SMS Filtering* agar cakupan perlindungan sistem menjadi lebih luas.
b. Mengimplementasikan fitur autentikasi dan *dashboard* pengguna untuk mencatat riwayat deteksi secara terpusat dalam *database*.
c. Membangun mekanisme *retraining* otomatis menggunakan data dari *database* guna meningkatkan akurasi algoritma secara berkelanjutan.

Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam upaya mitigasi ancaman siber, khususnya kejahatan *phishing* yang kian marak. Melalui pengembangan sistem yang mengombinasikan kekuatan kecerdasan buatan dan kontrol logika yang terstruktur, diharapkan pengguna dapat memiliki alat pertahanan yang lebih andal dalam menjaga keamanan data pribadi di ruang digital. Kesuksesan implementasi ini diharapkan menjadi fondasi bagi pengembangan sistem keamanan siber yang lebih adaptif, cerdas, dan komprehensif di masa depan.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih yang sebesar-besarnya kepada dosen pembimbing atas bimbingan dan arahannya dalam penyelesaian penelitian ini, serta kepada keluarga saya atas dukungan moral maupun materi yang tiada henti. Apresiasi juga diberikan kepada rekan-rekan sejawat yang telah memberikan bantuan teknis dan motivasi selama proses pengembangan sistem hingga jurnal ini dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

[1] E. Susanto, Lady Antira, K. Kevin, E. Stanzah, and A. A. Majid, "Manajemen Keamanan Cyber di Era Digital," *Journal of Business and Entrepreneurship*, vol. 11, no. 1, pp. 23–33, 2023, doi: 10.46273/job&e.v11i1.365.

[2] F. C. Y. Y. Pangalila, C. H. J. De Fretes, and R. O. C. Seba, "Peran National Central Bureau (NCB)-Interpol Indonesia dalam Penanganan Cybercrime (Romance Scam) Tahun 2018-2021," *Intermestic: Journal of International Studies*, vol. 8, no. 1, p. 356, Nov. 2023, doi: 10.24198/intermestic.v8n1.17.

[3] A. D. Harahap, D. Juardi, and A. S. Y. Irawan, "RANCANG BANGUN SISTEM PENDETEKSI LINK PHISHING MENGGUNAKAN ALGORITMA RANDOM FOREST BERBASIS WEB," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, pp. 2677–686, Aug. 2024, doi: 10.23960/jitet.v12i3.4858.

[4] A. Aman, "Pengujian Keamanan Jaringan Nirkabel Melalui Simulasi Serangan Man In The Middle Attack Di Sekolah XYZ," *Digital Transformation Technology*, vol. 3, no. 2, pp. 824–831, Dec. 2023, doi: 10.47709/digitech.v3i2.3378.

[5] A. Bichnigauri, I. Kartvelishvili, L. Shonia, D. Bichnigauri, and O. Gudadze, "Unveiling Quishing: The Dark Side of QR Codes in Cyber attacks," თავდაცვა და მეცნიერება, vol. 2, Dec. 2023, doi: 10.61446/ds.2.2023.7412.

[6] F. Sharevski, A. Devine, E. Pieroni, and P. Jachim, "Gone Quishing: A Field Study of Phishing with Malicious QR Codes," Apr. 2022, [Online]. Available: http://arxiv.org/abs/2204.04086

[7] M. Weinz, N. Zannone, L. Allodi, and G. Apruzzese, "The Impact of Emerging Phishing Threats: Assessing Quishing and LLM-generated Phishing Emails against Organizations," in *Proceedings of the ACM Conference on Computer and Communications Security*, Association for Computing Machinery, Aug. 2025, pp. 1550–1566. doi: 10.1145/3708821.3736195.

[8] S. Tandel, J. Chordiya, and P. S. H. Patil, "Tricked by the Square: Investigating the Rise and Reach of Quishing Attacks," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 13, no. 4, pp. 1823–1837, Apr. 2025, doi: 10.22214/ijraset.2025.68660.

[9] A. Prana Walidin *et al.*, "KALI LINUX SEBAGAI ALAT ANALISIS KEAMANAN JARINGAN MELALUI PENGGUNAAN NMAP, WIRESHARK, DAN METASPLOIT," Feb. 2025. doi: https://doi.org/10.36040/jati.v9i1.12661.

[10] S. N. Adzimi, H. A. Alfasih, F. N. G. Ramadhan, S. N. Neyman, and A. Setiawan, "Implementasi Konfigurasi Firewall dan Sistem Deteksi Intrusi menggunakan Debian," *Journal of Internet and Software Engineering*, vol. 1, no. 4, pp. 1–12, Jun. 2024, doi: 10.47134/pjise.v1i4.2681.

[11] R. Fauzan, A. V. Vitianingsih, D. Cahyono, A. L. Maukar, and Y. A. B. Suprio, "Penerapan Algoritma Klasifikasi pada Machine Learning untuk Deteksi Phishing," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 2, pp. 531–540, Mar. 2025, doi: 10.57152/malcom.v5i2.1968.

[12] A. F. Mahmud and S. Wirawan, "Deteksi Phishing Website menggunakan Machine Learning Metode Klasifikasi," Jul. 2024. doi: https://doi.org/10.32520/stmsi.v13i4.3456.

[13] A. Ferdita Nugraha, R. Faticha, A. Aziza, and Y. Pristyanto, "Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing," *Jurnal Infomedia : Teknik Informatika, Multimedia, dan Jaringan*, vol. 7, no. 1, pp. 39–44, Jun. 2022, doi: http://dx.doi.org/10.30811/jim.v7i1.2959.

[14] G. Sinulingga, "Kajian Penipuan Social Engineering dalam Logistik E Commerce: Studi Kasus QRIS dan Airwaybiil Paket di Indonesia," Dec. 2025. doi: 10.64694/jmrbi.v14i03.159.

[15] A. Pradana and S. Susanto, "Implementasi Model Machine Learning untuk Deteksi Phishing dengan Pendekatan Ekstraksi Fitur yang Dioptimalkan," *Jurnal Teknologi Informasi dan Multimedia*, vol. 8, no. 1, pp. 27–40, Jan. 2026, doi: 10.35746/jtim.v8i1.881.

[16] R. P. Erdiyanto, "PENIPUAN MENGATASNAMAKAN BANK BERBENTUK PHISING," 2023. [Online]. Available: https://jig.rivierapublishing.id/index.php/rv/index

[17] R. Diana, H. Warni, and T. Sutabri, "PENGGUNAAN TEKNOLOGI MACHINE LEARNING UNTUK PELAYANAN MONITORING KEGIATAN BELAJAR MENGAJAR PADA SMK BINA SRIWIJAYA PALEMBANG," *JUTEKIN (Jurnal Teknik Informatika)*, vol. 11, no. 1, Jun. 2023, doi: 10.51530/jutekin.v11i1.709.

[18] S. Mahmuda, "Implementasi Metode Random Forest pada Kategori Konten Kanal Youtube," *Jurnal Jendela Matematika*, vol. 2, 2024.

[19] M. Prima and P. Silitonga, "IMPLEMENTASI KODE QR PENGIRIMAN KARTU HASIL STUDI MAHASISW," 2018.