Indonesian Journal of Multidisciplinary on Social and Technology
Homepage: https://journal.ilmudata.co.id/index.php/ijmst
Vol. 5 No. 1 (2026) pp: 1081-1089
P-ISSN: 2986-6790, e-ISSN: 2986-6782

Rancang Bangun *Extension Browser* Berbasis *Heuristic Analysis* untuk Deteksi dan Pemblokiran *Phishing*

Umar Faruq Manek, Muh.Yamin, Bambang Pramono, L.M Fid Aksara, Natalis Ransi, Isnawaty
Jurusan Informatika, Fakultas Teknik, Universitas Halu Oleo
umarfaruqmanek@gmail.com, muh_yamin@uho.ac.id, 1bambang.pramono@uho.ac.id, fid.aksara@uho.ac.id, natalis.ransi@uho.ac.id, isnawaty@uho.ac.id

Abstrak

Serangan phishing memanfaatkan penyamaran alamat dan identitas situs untuk memperoleh informasi sensitif pengguna. Pemeriksaan yang hanya mengandalkan daftar URL berbahaya belum selalu mengenali situs baru, sedangkan klasifikasi berbasis URL dapat kehilangan konteks aktivitas pada halaman. Penelitian ini bertujuan merancang dan membangun extension browser yang mendeteksi serta memblokir situs phishing secara otomatis dengan menyajikan alasan pemeriksaan. Metode penelitian menggunakan pengembangan iterative incremental melalui pengumpulan data, perancangan, implementasi, dan pengujian. Sistem menggabungkan URL Analyzer berbasis algoritma XGBoost dengan 20 fitur URL, DOM Analyzer untuk memeriksa elemen autentikasi dan masukan sensitif, serta Brand Analyzer untuk menguji kesesuaian identitas brand dengan domain. Hasil ketiganya dihubungkan melalui 19 aturan korelasi dan kebijakan keputusan yang dijalankan pada sisi klien. Model dilatih pada 612.713 URL dan dievaluasi pada 65.402 URL uji dengan akurasi 93,20%. Pemeriksaan terhadap 526 URL dari arsip pengujian menghasilkan akurasi 92,02%, presisi 93,68%, recall 90,11%, dan skor F1 91,86%. Evaluasi keputusan akhir terhadap 738 pengamatan memperoleh ketepatan 88,48%, dengan ketepatan kelas legitimate 94,58% dan phishing 82,38%. Seluruh 40 kasus pengujian cabang lulus dan tiga halaman tiruan pada simulasi lokal berhasil diblokir. Hasil menunjukkan bahwa integrasi klasifikasi URL dan bukti halaman dapat mendukung peringatan serta pemblokiran yang dapat dijelaskan, meskipun kesalahan klasifikasi masih ditemukan pada kedua kelas.

Kata kunci: Extension Browser, Heuristic Analysis, XGBoost, Deteksi Phishing, Pemblokiran Situs

Abstract

Phishing attacks impersonate website addresses and identities to obtain sensitive user information. Detection based solely on known malicious URL lists may miss newly created websites, whereas URL-based classification alone may overlook the behavior and identity displayed by a webpage. This study aimed to design and develop a browser extension that automatically detected and blocked phishing websites while explaining its decisions. The research followed an iterative incremental development approach involving data collection, system design, implementation, and testing. The proposed system combined an XGBoost-based URL Analyzer using 20 URL features, a DOM Analyzer examining authentication and sensitive input elements, and a Brand Analyzer assessing the relationship between claimed brands and website domains. Their outputs were integrated using 19 correlation rules and a client-side decision policy. The model was trained using 612,713 URLs and evaluated on 65,402 test URLs, achieving 93.20% accuracy. An evaluation of 526 URLs from a previous testing archive yielded 92.02% accuracy, 93.68% precision, 90.11% recall, and a 91.86% F1-score. Final decisions for 738 observations achieved 88.48% overall accuracy, with class-specific correctness of 94.58% for legitimate websites and 82.38% for phishing websites. All 40 branch-testing cases passed, and three imitation pages were blocked in controlled local simulations. The findings indicate that combining URL classification with webpage evidence supports explainable warnings and blocking. However, false classifications still occur in both classes, and the archived evaluation should not be interpreted as a fully independent external test.

Keywords: Browser Extension, Heuristic Analysis, XGBoost, Phishing Detection, Website Blocking

1. Pendahuluan

Perluasan layanan digital meningkatkan frekuensi pengguna memasukkan identitas pribadi, kata sandi, dan informasi keuangan melalui situs web. Kondisi tersebut dimanfaatkan dalam *phishing*, yakni serangan yang menyamar sebagai layanan tepercaya untuk menipu pengguna. Data yang dikutip dalam skripsi menunjukkan APWG mengamati 1.003.924 serangan secara global pada kuartal I 2025, sedangkan BSSN mencatat 26.771.610 aktivitas terindikasi *phishing* di Indonesia sepanjang 2024. Angka tersebut menggunakan satuan pengukuran

berbeda, tetapi sama-sama menunjukkan pentingnya perlindungan ketika pengguna mengakses halaman web (Alwan et al., 2026).

Serangan tidak terbatas pada penggunaan *domain* yang sepenuhnya berbeda. Penyerang dapat memanfaatkan manipulasi *subdomain*, penggantian karakter pada nama layanan, peniruan tampilan, dan penggunaan HTTPS agar halaman terlihat wajar (Alwan et al., 2026). Oleh karena itu, keberadaan HTTPS atau tercatatnya suatu URL dalam daftar blokir tidak cukup sebagai satu-satunya dasar keputusan. Tinjauan Safi dan Singh (2023) menguraikan berbagai teknik deteksi situs *phishing*, sedangkan Pourmohamad et al. (2024) menunjukkan perlunya menilai mekanisme antiphishing pada sisi klien secara berkelanjutan.

Penelitian terdahulu telah menerapkan beberapa pendekatan. Al Ghifani et al. (2022) mengembangkan extension Google Chrome berbasis Random Forest yang dapat berjalan tanpa layanan web. Abdul et al. (2025) menggunakan XGBoost melalui layanan Flask; Siriwardhana et al. (2026) juga menghubungkan prediksi XGBoost berbasis URL dengan *extension browser*. Pendekatan berbasis fitur situs dan pembelajaran mesin dilaporkan oleh Shajahan et al. (2025), sedangkan Dandotiya et al. (2026) mengkaji deteksi melalui extension yang ditingkatkan dengan pembelajaran mesin. Studi lain mengembangkan sistem pemeriksaan URL, termasuk penggunaan support vector machine (Aryanti & Nabila, 2025), pengolahan bahasa alami (Nabila et al., 2025), dan deteksi *domain* pada *browser* (Senanayake et al., 2025).

Meskipun klasifikasi URL dapat digunakan sebagai pemeriksaan awal, hasil prediksi tidak selalu merepresentasikan tujuan formulir, klaim identitas, atau pola autentikasi pada halaman. Demikian pula keberadaan elemen *login* tidak secara otomatis berarti *phishing* karena layanan resmi juga menggunakan halaman autentikasi dan OAuth. Sifat kontekstual ini mendasari perlunya menghubungkan hasil klasifikasi URL dengan bukti dari halaman, bukan sekadar menjumlahkan skor risiko. Gagasan peringatan yang menampilkan indikator kontekstual juga dikaji oleh Roy et al. (2025), sementara Vignesh dan Anandan (2025) membahas pemeriksaan berbasis aturan melalui *extension browser*.

Berdasarkan kesenjangan tersebut, penelitian ini merancang dan membangun *extension browser* yang melakukan pemeriksaan melalui tiga modul: *URL Analyzer*, *DOM Analyzer*, dan *Brand Analyzer*. Prediksi XGBoost atas 20 fitur URL dikorelasikan dengan bukti struktur halaman dan kesesuaian identitas *brand* melalui 19 aturan serta kebijakan keputusan. Tujuan penelitian adalah menghasilkan deteksi dan pemblokiran situs *phishing* saat halaman diakses, sekaligus menampilkan dasar keputusan kepada pengguna. Kebaruan rancangan terletak pada pemisahan analisis URL berbasis model dari bukti DOM dan *brand* berbasis aturan, kemudian penggabungannya di sisi klien.

2. Metode Penelitian

Penelitian dilakukan dengan pendekatan iterative incremental melalui tahap perencanaan, analisis kebutuhan, perancangan, implementasi dan pengujian, serta evaluasi dan iterasi. Studi literatur digunakan untuk menentukan parameter deteksi URL, pola autentikasi dan pengumpulan informasi sensitif, serta identitas *brand*. Tahap berikutnya mencakup pengolahan data berlabel, pelatihan XGBoost, implementasi analisis berbasis aturan, integrasi kebijakan keputusan, dan pengujian sistem. Pendekatan tersebut memungkinkan pengujian setiap komponen sebelum integrasi keseluruhan.

Dataset utama berasal dari PhreshPhish yang menyediakan pasangan URL dan label *phishing* atau *legitimate*. Data tambahan dari kumpulan URL yang telah dihimpun sebelumnya digunakan untuk memperluas variasi. Pembersihan data mencakup validasi format, normalisasi skema protokol, pemeriksaan duplikasi, konflik label, dan ekstraksi fitur. Dari 666.315 URL awal PhreshPhish, sebanyak 18 data bermasalah dikeluarkan dan 5.741 URL *legitimate* tanpa protokol dinormalisasi menggunakan HTTPS. Setelah digabung dengan data tambahan dan melalui pengolahan akhir, kumpulan yang digunakan untuk pembentukan model berjumlah 678.115 URL.

Tabel 1. Komposisi Data URL untuk Pembentukan Model

Kelas	Jumlah URL
Phishing	289.907
Legitimate	388.208
Total	678.115

DOI: https://doi.org/10.69693/ijmst.v5i1.14671
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Data dibagi berdasarkan kelompok *domain* agar URL dari *domain* yang sama tidak tersebar pada data latih dan data uji. Sebanyak 612.713 URL dialokasikan untuk pelatihan dan 65.402 URL untuk evaluasi model, dengan proporsi masing-masing 90,36% dan 9,64%. Pembagian secara berkelompok digunakan untuk mengurangi kemiripan karakteristik *domain* antara kedua kumpulan data. Label *legitimate* dikodekan sebagai 0 dan *phishing* sebagai 1.

URL Analyzer menggunakan 20 fitur leksikal dan struktural, mencakup panjang URL, *hostname*, path dan query; jumlah *subdomain*, kedalaman path, titik dan tanda hubung; proporsi angka; karakter khusus; simbol @; penggunaan host IP dan HTTPS; entropi URL serta *hostname*; karakteristik token sensitif; panjang label *domain*; serta rangkaian angka pada *hostname*. *Domain age* dari RDAP/WHOIS dan extension nama berkas URL diperiksa sebagai informasi pendukung terpisah, bukan sebagai fitur masukan XGBoost. Ekstraksi untuk pelatihan menggunakan definisi yang sama dengan ekstraksi pada extension saat inferensi.

Model XGBoost dilatih menggunakan pustaka Python, dengan 500 pohon keputusan, kedalaman maksimum 7 dan learning rate 0,05. Model terlatih dikonversi agar inferensi dapat dijalankan melalui JavaScript di sisi klien. Probabilitas *phishing* lebih besar atau sama dengan 0,50 dipetakan ke kelas *phishing*; hasil di bawah ambang tersebut dipetakan ke kelas *legitimate*. Konfigurasi inti yang dicantumkan dalam skripsi diringkas pada Tabel 2.

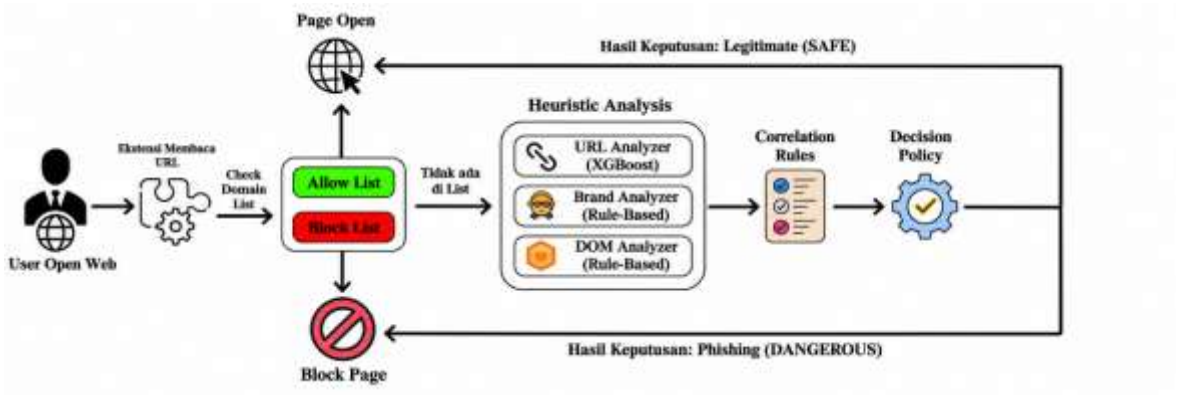
Tabel 2. Konfigurasi Utama Model XGBoost

Parameter	Nilai
Jumlah fitur	20
n_estimators	500
max_depth	7
learning_rate	0,05
min_child_weight	12
subsample	0,85
colsample_bytree	0,90
reg_lambda / reg_alpha	5 / 0,10
objective	binary:logistic
Ambang keputusan	0,50

Komponen *DOM Analyzer* membaca kolom kata sandi, isian informasi sensitif, tujuan formulir lintas *domain*, tombol *login*, dan pengalihan autentikasi; pemeriksaan tambahan mencakup tautan unduhan serta karakteristik tujuannya. *Brand Analyzer* mendeteksi klaim identitas pada judul, heading, logo, dan *hostname*, lalu membandingkannya dengan *domain* referensi yang diperoleh dari sumber pendukung. Data RDAP/WHOIS, Clearbit, Brandfetch, dan sumber *brand-domain* digunakan sebagai informasi tambahan; layanan tersebut tidak menjadi mesin klasifikasi utama.

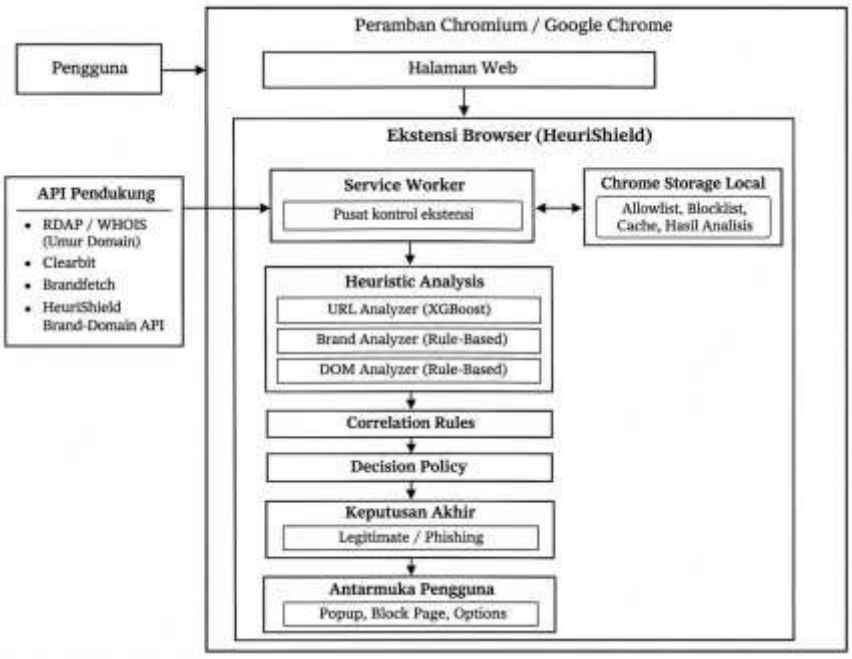
Setiap komponen menghasilkan bukti yang kemudian dievaluasi melalui 19 *correlation rules* (R01–R19) dan *decision policy*. Aturan mempertimbangkan hubungan antarindikator, termasuk konflik *brand-domain* bersama formulir sensitif, konteks OAuth resmi, serta informasi pendukung *domain*. Proses dimulai dari pemeriksaan *allowlist* dan *blocklist* lokal, dilanjutkan analisis otomatis jika *domain* belum tercantum pada daftar manual. Keputusan akhir terdiri atas *legitimate* dan *phishing*; apabila diputuskan *phishing*, pengguna diarahkan ke halaman blokir dengan alasan pemeriksaan.

DOI: https://doi.org/10.69693/ijmst.v5i1.14671
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 1. Alur Pemeriksaan dan Penentuan Keputusan Akhir Extension

Arsitektur implementasi memanfaatkan service worker sebagai pengendali, content script untuk membaca halaman, dan penyimpanan lokal untuk riwayat serta daftar *domain*. *URL Analyzer* menjalankan model XGBoost secara lokal, sedangkan DOM dan *brand* dianalisis menggunakan aturan. Gambar 2 memperlihatkan posisi komponen pada peramban serta hubungannya dengan layanan pendukung yang memerlukan jaringan.



Gambar 2. Arsitektur Extension Browser dan Layanan Pendukung

Tabel 3. Parameter Utama Pemeriksaan DOM dan Brand

Modul	Parameter	Aspek yang diperiksa
DOM Analyzer	has_password_field	Kolom kata sandi
DOM Analyzer	has_sensitive_input	Masukan informasi sensitif
DOM Analyzer	form_cross_domain	Tujuan formulir lintas domain
DOM Analyzer	login_redirect_detected	Pengalihan dalam proses autentikasi
Brand Analyzer	brand_in_title / brand_in_heading	Identitas brand pada judul atau heading
Brand Analyzer	official_domain_match	Kesesuaian domain dengan referensi resmi
Brand Analyzer	brand_domain_mismatch	Konflik identitas brand dan domain

DOI: https://doi.org/10.69693/ijmst.v5i1.14671
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Pengujian dilakukan pada empat tingkat. Pertama, model dievaluasi pada 65.402 URL yang disisihkan dari pelatihan. Kedua, klasifikasi *URL Analyzer* dievaluasi ulang pada arsip 526 URL yang tidak termasuk kumpulan pembentukan model, masing-masing 263 *phishing* dan *legitimate*; arsip tersebut pernah berperan dalam pengembangan sehingga tidak dianggap sebagai evaluasi independen baru. Ketiga, keputusan akhir diuji pada 738 pengamatan, terdiri atas 369 per kelas. Keempat, pengujian *white-box* dengan 40 kasus memeriksa percabangan logika, sedangkan tiga situs tiruan berbasis HTTPS dijalankan pada lingkungan lokal terkendali untuk memeriksa mekanisme pemblokiran.

Perhitungan evaluasi klasifikasi menggunakan *phishing* sebagai kelas positif. *True positive* menunjukkan URL *phishing* yang diprediksi *phishing*, sedangkan *true negative* menunjukkan URL *legitimate* yang dikenali dengan benar. *False positive* merepresentasikan URL *legitimate* yang dinyatakan *phishing* dan *false negative* merepresentasikan URL *phishing* yang luput dari deteksi. Akurasi digunakan untuk mengukur proporsi seluruh prediksi benar, presisi menunjukkan ketepatan prediksi *phishing*, *recall* mengukur proporsi *phishing* yang berhasil dikenali, dan F1-score digunakan untuk merangkum keseimbangan presisi serta *recall*. Seluruh metrik dihitung berdasarkan label acuan pada setiap himpunan uji.

Pada pengujian keputusan akhir, unit analisis adalah pengamatan hasil extension setelah tahap korelasi dan kebijakan keputusan, bukan sekadar keluaran probabilitas XGBoost. Karena himpunan 65.402, 526, dan 738 memiliki tujuan serta prosedur yang berlainan, metrik setiap tahap dilaporkan secara terpisah. Analisis pembahasan menitikberatkan pada jenis kesalahan kedua kelas, bukti kontekstual yang mengubah keputusan, dan keterbatasan validitas pengujian arsip maupun simulasi lokal.

3. Hasil dan Pembahasan

Implementasi menghasilkan extension yang melakukan pemeriksaan URL secara otomatis ketika halaman diakses. Pemeriksaan daftar izin dan daftar blokir manual ditempatkan sebelum inferensi model. Hasil *URL Analyzer* diperkaya dengan bukti elemen halaman dan identitas *brand*, kemudian diproses melalui aturan korelasi untuk menghasilkan keputusan akhir. Antarmuka utama menampilkan status halaman, *domain*, informasi identitas *brand*, umur *domain* jika tersedia, riwayat, serta tautan untuk meninjau alasan pemeriksaan.



Gambar 3. Panel Utama Extension yang Menampilkan Hasil Pemeriksaan

Proses analisis URL berjalan melalui model lokal dengan 20 fitur numerik. Pengujian pada 65.402 URL menghasilkan 60.952 prediksi benar dan 4.450 salah. Kesalahan tersebut meliputi 1.545 URL *legitimate* yang diprediksi *phishing* dan 2.905 URL *phishing* yang diprediksi *legitimate*. Nilai akurasi, presisi, *recall*, dan skor F1 pada model hasil pelatihan diperlihatkan pada Tabel 3.

Tabel 4. Evaluasi Model XGBoost pada Data Uji Terpisah

Data Uji	Akurasi	Presisi	Recall	F1-score
65.402	93,20%	94,44%	90,03%	92,18%

DOI: https://doi.org/10.69693/ijmst.v5i1.14671
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Metrik pada Tabel 4 menggambarkan performa model dalam klasifikasi URL dengan komposisi data uji yang berasal dari pemisahan berdasarkan *domain*. *Recall* 90,03% mengindikasikan bahwa sebagian URL *phishing* masih terlewat oleh pemeriksaan awal. Karena itu, rancangan memasukkan pemeriksaan identitas *brand* dan karakteristik formulir pada halaman sebagai dasar untuk mempertimbangkan kembali keputusan. Namun, pengujian model ini berbeda ruang lingkupnya dengan pengujian keputusan akhir seluruh extension.

Dalam implementasi aturan korelasi, nilai hasil analisis tidak dijumlahkan menjadi skor risiko tunggal. Aturan mencari hubungan spesifik antarbukti; sebagai contoh, identitas *brand* yang tidak sesuai dengan *domain* bersama temuan kolom kata sandi memperkuat indikasi peniruan kredensial. Aturan lain memeriksa konteks autentikasi OAuth resmi untuk mengurangi risiko halaman layanan yang sah diblokir hanya karena pola leksikal URL. Contoh aturan yang dipakai dalam implementasi disajikan pada Tabel 5.

Tabel 5. Contoh Aturan Korelasi yang Diimplementasikan

Aturan	Kondisi pokok	Peran terhadap keputusan
R03	Konflik brand-domain terkonfirmasi serta terdapat kolom kata sandi atau masukan sensitif	Bukti peniruan kredensial
R06	Kontrol OAuth, penyedia dikenal, dan navigasi menuju penyedia resmi	Bukti konteks autentikasi resmi
R09	Indikasi kuat peniruan brand pada hostname	Bukti peniruan identitas
R12	Prediksi URL phishing, peniruan brand yang kuat, dan tautan unduhan	Memperkuat keputusan phishing
R14–R19	Prediksi awal phishing memenuhi serangkaian syarat halaman resmi	Mendukung koreksi menjadi legitimate

Evaluasi ulang arsip pengujian *live* mencakup 526 URL, yaitu 263 *legitimate* dan 263 *phishing*. Sebanyak 484 URL diklasifikasikan dengan benar dan 42 salah. Berdasarkan confusion matrix, prediksi benar terhadap *phishing* berjumlah 237, prediksi benar terhadap *legitimate* 247, kesalahan positif 16, dan kesalahan negatif 26. Hasil pengujian tersebut diringkas pada Tabel 6.

Tabel 6. Confusion Matrix URL Analyzer pada Arsip 526 URL

Label acuan	Prediksi phishing	Prediksi legitimate
Phishing	237 (TP)	26 (FN)
Legitimate	16 (FP)	247 (TN)

Dari matriks tersebut, diperoleh akurasi 92,02%, presisi 93,68%, *recall* 90,11%, dan F1-score 91,86%. Sebanyak 16 URL *legitimate* ditandai *phishing*, sedangkan 26 URL *phishing* diklasifikasikan *legitimate*. Angka ini menunjukkan kemampuan model pada URL arsip yang berbeda dari data pembentukan model, tetapi tidak boleh dibaca sebagai uji independen sepenuhnya karena arsip tersebut sebelumnya pernah memengaruhi proses pengembangan. Kebutuhan evaluasi tambahan pada data yang sama sekali belum digunakan selama pengembangan tetap terbuka.

Evaluasi keputusan akhir extension dilakukan terhadap 738 pengamatan. Hasilnya menunjukkan keputusan benar sebanyak 349 dari 369 pengamatan *legitimate* serta 304 dari 369 pengamatan *phishing*. Secara keseluruhan, keputusan benar berjumlah 653, sedangkan keputusan salah 85. Rincian tingkat ketepatan tiap kelas diperlihatkan pada Tabel 7.

Tabel 7. Evaluasi Keputusan Akhir Extension

Kelas	Pengamatan	Benar	Salah	Ketepatan
Legitimate	369	349	20	94,58%
Phishing	369	304	65	82,38%
Total	738	653	85	88,48%

Tabel 7 memperlihatkan bahwa ketepatan kelas *legitimate* lebih tinggi daripada kelas *phishing*, masing-masing 94,58% dan 82,38%. Ketimpangan ini menunjukkan bahwa 65 pengamatan *phishing* belum dikenali pada

DOI: https://doi.org/10.69693/ijmst.v5i1.14671
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

keputusan akhir. Hasil tersebut perlu dibaca sebagai evaluasi sistem terintegrasi, bukan bukti bahwa kombinasi modul secara umum lebih akurat daripada model URL saja, karena komposisi dan prosedur pengujian berbeda. Peningkatan performa pada skenario tertentu tidak serta-merta menghasilkan peningkatan metrik keseluruhan.

Pengujian *white-box* dengan metode *branch testing* mencakup WB-01 hingga WB-40. Cabang yang diperiksa meliputi masukan URL HTTP/HTTPS, pembacaan identitas tab, pemeriksaan *allowlist* dan *blocklist*, evaluasi R01–R19, serta prioritas *decision policy* dan penanganan kondisi informasi tidak tersedia. Seluruh 40 kasus menghasilkan keluaran sesuai dengan yang diharapkan. Temuan tersebut memvalidasi kesesuaian percabangan dengan rancangan pada lingkup kasus uji, tetapi tidak menghilangkan kemungkinan kesalahan klasifikasi ketika sistem menerima situs baru yang belum tercakup oleh skenario.

Di sisi pengguna, keputusan *phishing* menghasilkan halaman blokir yang menyertakan ringkasan dan alasan pemeriksaan, bukan hanya label berbahaya. Halaman ini dirancang untuk membantu pengguna memahami indikator yang ditemukan, termasuk konflik identitas *brand*, pola URL, atau formulir berisiko. Contoh keluaran halaman blokir hasil implementasi ditampilkan pada Gambar 4.

Fitur riwayat analisis menyimpan catatan pemeriksaan secara lokal dan memungkinkan pengguna meninjau URL, *domain*, keputusan, alasan, serta waktu pencatatan. Antarmuka menyediakan pencarian dan penyaringan yang bekerja pada catatan yang tersimpan. Fitur ini mendukung pemeriksaan ulang keputusan yang pernah dihasilkan, termasuk apabila pengguna perlu mengidentifikasi URL yang sebelumnya dinilai mencurigakan. Penyimpanan lokal digunakan untuk riwayat, daftar *domain* manual, dan pengaturan dasar, sedangkan akses ke layanan registrasi *domain* serta referensi identitas *brand* tetap dilakukan hanya ketika informasi pendukung dibutuhkan.

Pengaturan *allowlist* dan *blocklist* memberikan pengecualian manual terhadap hasil pemeriksaan otomatis. Pengguna dapat menambah atau menghapus *domain* pada daftar izin maupun daftar blokir. Ketika *domain* berada pada *allowlist*, analisis otomatis dilewati dan halaman diizinkan; ketika *domain* terdapat pada *blocklist*, halaman diblokir berdasarkan keputusan manual. Mekanisme ini memberikan kontrol kepada pengguna, tetapi pilihan memasukkan *domain* pada *allowlist* harus dipertimbangkan secara hati-hati karena dapat melewati pemeriksaan otomatis.

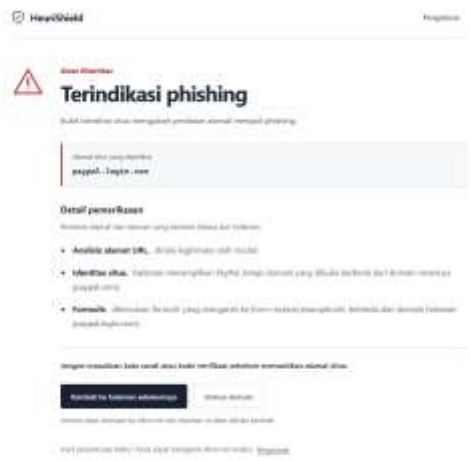


Gambar 4. Halaman Blokir dengan Informasi Alasan Pemeriksaan

Simulasi lokal menggunakan tiga halaman tiruan yang menyerupai Instagram, Facebook, dan PayPal dengan URL yang menyerupai alamat asli serta penggunaan HTTPS. Pada skenario Instagram, URL dan bukti formulir kredensial mendukung keputusan *phishing*. Pada skenario Facebook, *domain* yang menyerupai nama layanan dan kolom kata sandi menghasilkan pemblokiran meskipun tujuan formulir tetap pada *domain* yang sama. Pada skenario PayPal, *URL Analyzer* justru memberikan klasifikasi awal *legitimate*, tetapi analisis *brand* dan DOM

DOI: https://doi.org/10.69693/ijmst.v5i1.14671
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

menemukan ketidaksesuaian identitas dengan *domain* serta formulir sensitif lintas *domain* sehingga keputusan akhir menjadi *phishing*. Ketiga halaman tiruan berhasil diblokir dalam lingkungan terkontrol.



Gambar 5. Pemblokiran Halaman Tiruan PayPal pada Simulasi Lokal

Gambar 5 menunjukkan kasus koreksi keputusan setelah pemeriksaan halaman. Hal ini memberikan ilustrasi operasional atas fungsi aturan korelasi, khususnya ketika sinyal URL tidak cukup kuat. Hasil tersebut sejalan dengan gagasan bahwa kombinasi karakteristik URL, konteks halaman, dan identitas layanan dapat memperkaya dasar deteksi (Shajahan et al., 2025; Syahli et al., 2025). Namun, simulasi hanya mencakup tiga situs tiruan lokal sehingga tidak dapat dianggap sebagai pengukuran keandalan terhadap semua variasi serangan nyata.

Penerapan inferensi pada sisi klien memungkinkan keputusan klasifikasi utama dilakukan tanpa ketergantungan pada backend prediksi. Penggunaan layanan RDAP/WHOIS dan referensi *brand* tetap memerlukan akses jaringan untuk memperoleh informasi pendukung. Pembatasan ini menjadi konsekuensi desain: model lokal mendukung pemrosesan awal yang segera dilakukan saat URL diakses, tetapi ketepatan pemeriksaan konteks dapat berubah ketika data eksternal tidak tersedia. Kajian keamanan extension modern juga menekankan pentingnya memperhatikan relasi antara izin, privasi, dan kemampuan pemeriksaan (Polčák et al., 2025).

Keterbatasan yang teridentifikasi mencakup *false positive* pada sebagian situs *legitimate*, *false negative* pada situs *phishing* tertentu, ketergantungan sebagian bukti terhadap ketersediaan identitas *brand*, serta cakupan pengujian simulasi yang terbatas. Karakteristik situs dapat berubah menurut waktu, *domain*, dan mekanisme autentikasi. Karena itu, hasil pada 526 URL arsip dan 738 pengamatan perlu dilengkapi penelitian lanjutan dengan kumpulan data yang benar-benar independen, variasi situs yang lebih luas, dan pengukuran operasional. Kebutuhan evaluasi pada data bervariasi juga didukung oleh tinjauan mengenai teknik dan *dataset* URL berbahaya (Tian et al., 2025) serta penelitian pemeriksaan URL berbasis heuristik di sisi klien (Utikar et al., 2026).

4. Kesimpulan

Penelitian menghasilkan *extension browser* berbasis model XGBoost dan aturan heuristik untuk mendeteksi serta memblokir situs *phishing* dengan alasan pemeriksaan. XGBoost pada 65.402 URL uji mencapai akurasi 93,20%, sedangkan evaluasi arsip 526 URL mencapai 92,02%. Evaluasi keputusan akhir pada 738 pengamatan memperoleh ketepatan 88,48%. Seluruh 40 kasus *branch testing* sesuai hasil yang diharapkan dan tiga halaman tiruan berhasil diblokir. Keterbatasan utama adalah kesalahan pada kedua kelas dan cakupan pengujian. Penelitian berikutnya disarankan memperluas data independen, memperbaiki verifikasi *brand*, dan menguji beragam skenario autentikasi untuk mengurangi kesalahan keputusan.

Referensi

1. Abdul, H., Afandi, K., Studi, P., Kriptografi, R., Siber, P., Studi, P., Kriptografi, R., Siber, P., Siber, B., Siber, B., Al-dzaki, M. L. F., Qomariasih, N., & Aulia, R. (2025). GuardSurfing: extension Browser sebagai Alat Bantu Deteksi Website Phishing dengan Metode Klasifikasi XGBoost untuk Deteksi URL Phishing Berbasis Flask Framework. Info Kripto, 19(2), 73–85.

DOI: https://doi.org/10.69693/ijmst.v5i1.14671
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

2. Al Ghifani, M. A. G., Hananto, B., & Wahyono, B. T. (2022). Implementasi extension Google Chrome Dalam Mendeteksi Situs Web Phishing Menggunakan Algoritma Random Forest. Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA), 3, 179–189. https://conference.upnvj.ac.id/index.php/senamika/article/view/2152
3. Alwan, F., Al Pasha, R., Kaspandiri, E. P., Dzaky, M., & Fariz, A. (2026). Analisis Serangan Phising dan Upaya Pencegahannya pada Sistem Informasi. Jurnal Ilmu Komputer, 2(2), 161–165.
4. APWG. (2025). Phishing Activity Trends Report.
5. Aryanti, A., & Nabila, N. (2025). Deteksi URL Phishing Menggunakan Algoritma Support Vector Machine Berbasis Website. Jurnal Resistor, 8(2), 46–58. https://ejournal.instiki.ac.id/index.php/jurnalresistor/article/view/1943
6. Badan Siber dan Sandi Negara. (2024). Lanskap Keamanan Siber Indonesia. ID-SIRTII/CC, 70, 1–107. bit.ly/44bzpHM
7. Dandotiya, M., Goyal, N. K., Khunteta, A., & Tiwari, B. (2026). Real time identification of phishing attacks through machine learning enhanced browser extensions. Scientific Reports, 16(1), 1–25. https://doi.org/10.1038/s41598-026-35655-7
8. Nabila, N., Hesti, E., & Aryanti, A. (2025). Deteksi URL Phishing Menggunakan Natural Language Processing dan Support Vector Machine Berbasis Machine Learning. Building of Informatics, Technology and Science (BITS), 7(1), 526–537. https://doi.org/10.47065/bits.v7i1.7443
9. Polčák, L., Maone, G., McMahon, M., & Bednář, M. (2025). Developers' Insight on Manifest v3 Privacy and Security Webextensions. International Conference on Web Information Systems and Technologies, WEBIST - Proceedings, 15–26. https://doi.org/10.5220/0013669000003985
10. Pourmohamad, R., Wirsz, S., Oest, A., Bao, T., Shoshitaishvili, Y., Wang, R., Doupé, A., & Bazzi, R. A. (2024). Deep Dive into Client-Side Anti-Phishing: A Longitudinal Study Bridging Academia and Industry. ACM AsiaCCS 2024 - Proceedings of the 19th ACM Asia Conference on Computer and Communications Security. https://doi.org/10.1145/3634737.3657027
11. Roy, S. S., Torres, C., & Nilizadeh, S. (2025). "Explain, Don't Just Warn!" — A Real-Time Framework for Generating Phishing Warnings with Contextual Cues. University of Texas at Arlington. http://arxiv.org/abs/2505.06836
12. Safi, A., & Singh, S. (2023). A systematic literature review on phishing website detection techniques. Journal of King Saud University - Computer and Information Sciences, 35(2), 590–611. https://doi.org/10.1016/j.jksuci.2023.01.004
13. Senanayake, J., Rajapaksha, S., Yanai, N., Kalutarage, H., & Komiya, C. (2025). MADONNA: Browser-based malicious domain detection using Optimized Neural Network by leveraging AI and feature analysis. Elsevier, 152, 104371. https://doi.org/10.1016/j.cose.2025.104371
14. Shajahan, S., George, T., V, J. P., M, S. K., Krishna, S., Krishna, S., & Jagadish, N. (2025). Phishing Website Detection Using Machine Learning. International Journal of Research Publication and Reviews, 6(3), 10039–10043. https://doi.org/10.55248/gengpi.6.0325.1325
15. Siriwardhana, K. U. T., Sharfana, A. F., & Rathnayaka, R. M. N. (2026). Real-Time URL-Based Phishing Detection Using XGBoost and Browser Extension Integration. Journal of Information and Communication Technology, 03(02), 55–64.
16. Syahli, A., Kurniati, R., & Hidayasari, N. (2025). Alat OSINT Berbasis Web untuk Deteksi URL Phishing Menggunakan Integrasi API. Techno.Com, 24(3), 905–918. https://doi.org/10.62411/tc.v24i3.13769
17. Tian, Y., Yu, Y., Sun, J., & Wang, Y. (2025). From past to present: A survey of malicious URL detection techniques, datasets and code repositories. Computer Science Review, 58, 1–37. https://doi.org/10.1016/j.cosrev.2025.100810
18. Utikar, A., Kadam, A., & Bhujbal, T. (2026). Phish Guard: A Client-Side Heuristic-Based Phishing URL Detection System. International Journal of Innovative Research in Technology, 12(12), 3743–3750.
19. Vignesh, & Anandan. (2025). Rule-Based Approach for Detecting Phishing Websites through Browser Extension. International Journal of Scientific Research in Engineering and Management, 09(10), 1–9. https://doi.org/10.55041/ijsrem53386

DOI: https://doi.org/10.69693/ijmst.v5i1.14671
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)