Deteksi SEO-Spam Judi pada Domain Akademik menggunakan Fitur Struktural dan Tekstual dengan Evaluasi *Domain-Holdout*

Detecting Gambling SEO Spam on Academic Domains using Structural and Textual Features: A Domain-Holdout Evaluation

[1]Muh Ghazy Daffa Sampe, [2]Muhammad Kevin Adli Pratama, [3]Muhamad Rayhan Akhsani Taqwim, [4]Kalingga Dwindra Putraka, [5]Anindita Septiarini, [6]Novianti Puspitasari*
[1,2,3,4,6]Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman
[5]Magister Informatika, Fakultas Teknik, Universitas Mulawarman
[1,2,3,4,5,6]Jl. Sambaliung, Kota Samarinda, Indonesia
*e-mail: novipuspitasari@unmul.ac.id

(***received*:** 20 May 2026, ***revised*:** 4 June 2026, ***accepted*:** 28 July 2026)

Abstrak

Praktik spam SEO yang berkaitan dengan perjudian di domain akademis Indonesia menimbulkan risiko ganda: hal ini merusak integritas situs web institusi dan menimbulkan beban operasional bagi tim keamanan yang harus membedakan halaman yang disusupi dari konten yang sah. Penelitian ini mengembangkan pendekatan deteksi terawasi untuk mengidentifikasi halaman spam SEO yang terkait dengan perjudian di domain .ac.id dengan menggunakan kumpulan data acuan yang dikurasi secara ketat. Kumpulan data akhir berisi 2.003 halaman yang diberi anotasi secara manual, yang dikumpulkan dari 938 domain unik, dengan 815 halaman berbahaya dan 1.188 halaman yang tidak berbahaya. Pipeline yang diusulkan menggabungkan sinyal tekstual dari TF-IDF tingkat karakter dengan indikator struktural halaman seperti elemen tersembunyi, tautan mencurigakan, iframe, dan pola tautan eksternal. Untuk menghindari perkiraan yang terlalu optimis, evaluasi dilakukan dengan protokol domain-holdout di mana 20% domain dikecualikan sepenuhnya dari pelatihan dan pemilihan model. Lima model dibandingkan: *baseline* berbasis aturan kata kunci, *Naive Bayes*dengan TF-IDF kata, *Random Forest* dengan fitur struktural, Regresi Logistik hibrida, dan *SVM* hibrida. Hasil eksperimen pada set holdout menunjukkan bahwa *Random Forest* dengan fitur struktural mencapai skor F1 tertinggi sebesar 0,8844, sementara *SVM* hibrida yang diusulkan mencapai presisi tertinggi sebesar 0,9205 dengan skor F1 sebesar 0,8663. Temuan ini menunjukkan bahwa sinyal kompromi struktural lebih tangguh daripada petunjuk tekstual dalam skenario spam SEO terselubung, sementara model hibrida tetap menarik ketika presisi tinggi dan keterbacaan diprioritaskan.

Kata kunci: deteksi halaman berbahaya, domain akademik, domain-holdout, machine learning, SEO-spam judi.

Abstract

SEO spam practices related to gambling on Indonesian academic domains pose dual risks: they undermine the integrity of institutional websites and impose an operational burden on security teams tasked with distinguishing compromised pages from legitimate content. This study develops a supervised detection approach to identify gambling-related SEO spam pages on .ac.id domains using a rigorously curated reference dataset. The final dataset contains 2,003 manually annotated pages collected from 938 unique domains, comprising 815 malicious pages and 1,188 benign pages. The proposed pipeline combines textual signals from character-level TF-IDF with structural page indicators, including hidden elements, suspicious links, iframes, and external-link patterns. To avoid overly optimistic performance estimates, evaluation was conducted using a domain-holdout protocol, in which 20% of the domains were completely excluded from model training and selection. Five models were compared: a keyword-based baseline, Naïve Bayes with word-level TF-IDF, Random Forest with structural features, and hybrid Logistic Regression and Support Vector Machine (SVM) models. Experimental results on the holdout set show that Random Forest with structural features achieved the highest F1-score of 0.8844, whereas the proposed hybrid SVM achieved the highest

 ISSN:2302-8149
e-ISSN:2540-9719

precision of 0.9205, with an F1-score of 0.8663. These findings indicate that structural compromise signals are more robust than textual cues in detecting stealthy SEO spam scenarios, while hybrid models remain promising when high precision and interpretability are prioritized.
Keywords: malicious page detection, academic domains, domains holdout, machine learning, gambling SEO spam.

1 Pendahuluan

Fenomena penyalahgunaan situs web akademik untuk menanam halaman promosi judi online atau SEO-spam judi semakin sering dijumpai pada domain perguruan tinggi Indonesia. Halaman semacam ini umumnya memanfaatkan subfolder, direktori usang, atau endpoint yang kurang terawat untuk menanamkan konten yang ditujukan ke mesin pencari. Masalahnya tidak hanya berkaitan dengan reputasi institusi, tetapi juga mengganggu proses pemeliharaan situs, meningkatkan risiko blacklist, dan membebani analis keamanan yang harus memilah antara halaman normal, spam SEO, dan indikasi kompromi yang lebih serius. Literatur mutakhir tentang phishing website detection dan *malicious* URL analysis menunjukkan bahwa halaman berbahaya dirancang untuk mengeksploitasi sinyal lexical, HTML, dan perilaku halaman secara bersamaan, sehingga karakternya sering berbeda dari dokumen web biasa [1], [2].

Pendekatan deteksi berbasis kata kunci saja cenderung rapuh pada kasus cloaked SEO-spam judi. Penyerang dapat menyamarkan isi halaman, memecah kata kunci, memindahkan muatan spam ke anchor text, atau memanfaatkan struktur HTML yang tidak wajar. Karena itu, studi terbaru tentang phishing dan *malicious* webpage detection menekankan bahwa sinyal struktural seperti pola tautan, jumlah elemen tersembunyi, serta ketidakwajaran link keluar dapat berperan penting dalam klasifikasi [2], [3]. Di sisi lain, representasi teks tetap penting karena halaman judi online biasanya meninggalkan jejak istilah seperti slot, rtp, jackpot, toto, atau variasi turunannya pada judul, snippet, dan anchor text. Kombinasi keduanya menjadikan tugas ini cocok ditangani sebagai masalah klasifikasi biner berbasis fitur heterogen.

Dalam beberapa studi terkini, representasi tekstual masih efektif untuk mendeteksi URL atau halaman phishing, terutama bila dipadukan dengan fitur URL dan HTML yang relevan [4], [5]. Namun, ulasan mutakhir juga menunjukkan bahwa performa terbaik pada tugas web-threat detection semakin sering dicapai oleh pendekatan multimodal yang menggabungkan sinyal lexical, struktural, dan konteks teknis lainnya [2], [8]. Karena itu, penelitian ini tidak hanya memodelkan teks, tetapi juga menggabungkan fitur struktural yang diekstrak dari ringkasan HTML ke dalam satu pipeline klasifikasi.

Masalah penting yang sering diabaikan pada riset klasifikasi halaman web adalah kebocoran evaluasi. Jika data dibagi secara acak, halaman dari domain yang sama dapat muncul di train dan test secara bersamaan. Padahal, pola struktur HTML, template tema, dan format URL pada satu domain cenderung homogen. Akibatnya, model dapat terlihat sangat akurat karena mempelajari karakteristik domain, bukan benar-benar menggeneralisasi ke domain baru. Pada *dataset* yang bersifat web-centric, protokol evaluasi berbasis domain jauh lebih relevan daripada *random split* biasa.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan menyusun dan mengevaluasi sistem deteksi SEO-spam judi pada domain akademik Indonesia dengan tiga kontribusi utama. Pertama, penelitian ini menyajikan benchmark berlabel manual berisi 2.003 halaman dari 938 domain .ac.id. Kedua, penelitian ini menggunakan protokol domain-holdout sehingga klaim performa tidak terdistorsi oleh kebocoran antar-domain. Ketiga, penelitian ini membandingkan *baseline* keyword, text-only, structural-only, dan model *hybrid* agar dapat diketahui modalitas fitur mana yang paling relevan untuk tugas deteksi ini. Hasil penelitian menunjukkan bahwa fitur struktural merupakan sinyal paling kuat pada *dataset* ini, sedangkan model *hybrid* menawarkan *precision* tertinggi dan tetap mudah diinterpretasikan untuk kebutuhan operasional.

2 Tinjauan Literatur

Ancaman keamanan siber pada website terus berkembang, ditandai dengan meningkatnya praktik penyalahgunaan domain yang sah untuk keperluan SEO-spam judi dan penyebaran konten berbahaya. Salah satu modus yang kerap dijumpai adalah penyusupan halaman bertema judi online ke dalam

domain akademik guna memanipulasi peringkat mesin pencari. Tang dan Mahmoud [1] mengungkapkan bahwa website berbahaya masa kini tidak hanya mengandalkan konten teks, melainkan juga mengeksploitasi karakteristik URL, struktur HTML, serta perilaku halaman untuk mengelabui sistem deteksi otomatis. Kondisi ini menjadikan pendekatan berbasis machine learning sebagai solusi yang semakin banyak diadopsi dalam upaya mendeteksi halaman berbahaya.

Berbagai penelitian seputar deteksi URL berbahaya dan phishing website membuktikan bahwa penggabungan fitur tekstual dan struktural dapat meningkatkan kualitas klasifikasi secara signifikan. Tian dkk. [2] menyebutkan bahwa pola tautan, iframe, elemen tersembunyi, dan struktur HTML yang tidak wajar merupakan indikator kunci dalam mengidentifikasi halaman berbahaya. Hal senada disampaikan oleh Sánchez-Paniagua dkk. [3], yang menemukan bahwa fitur teknologi web dan skrip mencurigakan berkontribusi besar terhadap akurasi deteksi phishing. Sementara itu, Jha dkk. [4] menunjukkan bahwa representasi TF-IDF yang dikombinasikan dengan algoritma Support Vector Machine (SVM) dan *Random Forest* mampu menghasilkan performa tinggi dalam klasifikasi website berbahaya.

Perkembangan terkini dalam bidang ini menunjukkan bahwa pendekatan *hybrid* terbukti semakin efektif menghadapi ancaman web yang terus berevolusi. Kavya dan Sumathi [8] menjelaskan bahwa metode deteksi phishing modern umumnya memadukan fitur leksikal, fitur struktural HTML, dan sinyal kontekstual untuk memperkuat ketahanan model terhadap teknik cloaking dan obfuscation. Di sisi lain, Mia dkk. [10] menekankan bahwa evaluasi dengan *random split* cenderung menghasilkan performa yang terlalu optimistis akibat kebocoran data antar-domain. Oleh sebab itu, protokol domain-holdout dinilai lebih representatif dalam mengukur kemampuan model untuk melakukan generalisasi pada domain yang belum pernah ditemui sebelumnya.

Mengacu pada kajian-kajian tersebut, sebagian besar penelitian yang ada masih menitikberatkan pada phishing website dan URL berbahaya secara umum, sedangkan penelitian khusus mengenai deteksi SEO-spam judi pada domain akademik di Indonesia masih sangat minim. Atas dasar itu, penelitian ini merancang sistem deteksi SEO-spam judi yang mengombinasikan fitur tekstual dan struktural dengan menerapkan evaluasi domain-holdout, sehingga diharapkan mampu menghasilkan performa klasifikasi yang lebih realistis dan relevan dalam konteks lingkungan website akademik.

3 Metode Penelitian

3.1 *Dataset* dan anotasi

Dataset utama pada penelitian ini adalah file yang berisi 2.003 sampel halaman web hasil scraping domain akademik Indonesia (Tabel 1). Setiap sampel terdiri atas URL sumber, judul halaman, potongan isi teks, serta sejumlah fitur struktural yang diekstrak dari HTML. Setelah proses pembersihan akhir, *dataset* memuat 815 halaman *malicious* dan 1.188 halaman benign yang berasal dari 938 domain unik. Semua label pada *dataset* final ditetapkan secara manual oleh satu anotator, sehingga file raw berlabel heuristik dari fase sebelumnya tidak digunakan dalam evaluasi paper.

Proses anotasi dilakukan untuk membedakan halaman yang benar-benar mengandung indikasi SEO-spam judi dari halaman normal. Sampel diberi label *malicious* apabila halaman mengandung promosi judi online, backlink atau anchor yang mengarah ke ekosistem judi online, pola istilah yang khas untuk trafik pencarian judi online, atau kombinasi sinyal teks dan struktur yang mengarah pada kompromi halaman. Sebaliknya, halaman diberi label benign bila konten merupakan materi akademik, jurnal, repositori, berita kampus, atau halaman umum lain yang tidak menunjukkan indikasi spam judi. Catatan anotasi dan *Review* notes dipertahankan pada *dataset* untuk membantu audit kesalahan dan interpretasi hasil.

 ISSN:2302-8149
e-ISSN:2540-9719

Tabel 1 Ringkasan *dataset* final

Komponen	Nilai
Total sampel	2.003
Domain unik	938
Label *malicious*	815 (40,7%)
Label benign	1.188 (59,3%)
Train domain	750 domain / 1.677 sampel
Test domain	188 domain / 326 sampel
Distribusi label test	99 *malicious* / 227 benign
Sumber label	Anotasi manual penuh pada *dataset* final

3.2 Representasi fitur dan model

Fitur yang digunakan dibagi menjadi dua kelompok. Kelompok pertama adalah fitur tekstual yang dibangun dari penggabungan page_title dan body_text_snippet, lalu diubah menjadi representasi TF-IDF. Untuk *baseline Naive Bayes*digunakan word TF-IDF standar. Untuk model *hybrid* digunakan character-level TF-IDF dengan analyzer char_wb dan rentang n-gram 3-5. Representasi karakter dipilih karena lebih tahan terhadap variasi ejaan, token terpotong, serta kata kunci manipulatif yang sering muncul pada halaman spam.

Kelompok kedua adalah fitur struktural yang mencerminkan pola kompromi halaman. Penelitian ini menggunakan sembilan fitur, yaitu total_links, external_links, hidden_divs, iframes, display_none_links, suspicious_scripts, suspicious_links, suspicious_urls, dan suspicious_anchors. Semua fitur numerik dibersihkan, diubah ke tipe numerik, lalu ditransformasikan dengan log1p untuk menstabilkan sebaran. Pada model hybrid, fitur struktural distandardisasi sebelum digabungkan dengan fitur teks. Dari sisi konseptual, fitur-fitur ini merepresentasikan intensitas tautan keluar, keberadaan elemen tersembunyi, injeksi skrip yang mencurigakan, serta kepadatan anchor yang berpotensi digunakan untuk SEO-spam judi; pola-pola semacam ini juga sering dilaporkan penting pada studi malicious URL dan phishing website detection modern [2], [3], [6]. Rincian lebih lanjut mengenai pembagian kelompok fitur utama beserta makna operasionalnya dapat dilihat pada Tabel 2.

Tabel 2 Kelompok fitur utama yang digunakan dalam penelitian

Kelompok	Fitur	Makna operasional
Teks	page_title, body_text_snippet	Mewakili kata, frasa, dan pola karakter yang berasosiasi dengan SEO-spam judi online
Struktural	total_links, external_links	Mengukur kepadatan tautan dan dominasi link ke luar domain
Struktural	hidden_*div*s, iframes, display_none_links	Menggambarkan elemen tersembunyi atau embedding yang sering dipakai untuk injeksi konten
Struktural	suspicious_scripts	Mewakili keberadaan skrip yang tidak lazim pada halaman normal
Struktural	suspicious_links, suspicious_urls, suspicious_anchors	Menangkap pola anchor dan URL yang mengarah ke konten judi atau backlink manipulatif

3.3 Protokol evaluasi

Evaluasi dilakukan dalam dua tahap. Tahap pertama adalah model *selection* menggunakan *StratifiedGroupKFold* berbasis base_domain. Skema ini memastikan bahwa semua halaman dari domain yang sama tetap berada pada fold yang sama. Tahap kedua adalah evaluasi utama paper menggunakan domain-holdout 80:20. Sebanyak 188 domain dikeluarkan sepenuhnya dari training dan hanya dipakai untuk test akhir. Dengan rancangan ini, hasil pengujian mencerminkan kemampuan model menghadapi domain baru yang belum pernah terlihat sebelumnya.

Metrik utama yang digunakan adalah *F1-score* karena tugas ini menuntut keseimbangan antara *precision* dan *recall*. *Precision* penting agar sistem tidak terlalu banyak memicu *false* alarm pada situs akademik yang sah, sedangkan *recall* penting agar halaman kompromi tidak terlewat. Selain F1, penelitian ini juga melaporkan *accuracy*, precision, *recall*, serta confidence interval bootstrap 95% pada test set. Praktik pelaporan multi-metrik semacam ini juga sejalan dengan pedoman pelaporan studi machine learning yang menekankan transparansi evaluasi dan pelaporan lengkap atas performa model [9]. Untuk model *hybrid* SVM, probabilitas hasil kalibrasi juga digunakan untuk mengevaluasi titik operasi *threshold* 0,85 yang dipertimbangkan untuk penggunaan operasional relevan dalam studi deteksi ancaman web terkini [11], [12].

4 Hasil dan Pembahasan

4.1 Hasil evaluasi utama

Hasil utama penelitian diringkas pada Tabel 3. Secara numerik, *Random Forest* dengan fitur struktural menghasilkan *F1-score* tertinggi, yaitu 0,8844. Akan tetapi, *hybrid SVM* mencapai *precision* tertinggi, yaitu 0,9205, dengan *F1-score* 0,8663. *Hybrid* Logistic Regression memiliki F1 yang hampir identik dengan *hybrid* SVM, sedangkan *baseline* rule-based menunjukkan *recall* sempurna tetapi *precision* jauh lebih rendah. *Naive Bayes*berbasis word TF-IDF menjadi model dengan performa paling rendah pada test set, yang mengindikasikan bahwa representasi teks semata tidak cukup untuk menangani kasus spam yang disamarkan.

Temuan ini penting karena menunjukkan bahwa model proposed tidak harus selalu menjadi model dengan F1 tertinggi untuk tetap bernilai secara ilmiah. Dalam konteks operasi keamanan, *precision* tinggi sering lebih diutamakan agar analis tidak dibanjiri alert yang keliru. Karena itu, penelitian ini menempatkan *Random Forest structural-only* sebagai *baseline* terbaik dari sisi F1, sedangkan *hybrid SVM* dipertahankan sebagai model proposed karena memberikan kombinasi *precision* tinggi, interpretabilitas koefisien linear, dan fleksibilitas *threshold* berbasis probabilitas.

Tabel 3 Hasil evaluasi pada domain-*holdout test set*

Model	F1	*Accuracy*	Precision	*Recall*	CI 95%
Rule-Based (Keyword)	0,8534	0,8957	0,7444	1,0000	[0,796; 0,899]
NB (Word-TF-IDF)	0,6706	0,8282	0,8028	0,5758	[0,577; 0,752]
RF (Struct Only)	0,8844	0,9294	0,8800	0,8889	[0,835; 0,927]
Hybrid SVM (proposed)	0,8663	0,9233	0,9205	0,8182	[0,816; 0,914]
Hybrid LogReg	0,8660	0,9202	0,8842	0,8485	[0,802; 0,911]

4.2 Temuan utama: sinyal struktural lebih dominan daripada teks

Salah satu temuan paling penting dari penelitian ini adalah dominasi fitur struktural. Pada evaluasi model *selection* berbasis *StratifiedGroupKFold*, *baseline structural-only* memperoleh rerata F1 sebesar 0,8605, jauh di atas model *text-only* sebesar 0,7794. Setelah diuji pada domain-holdout, pola tersebut tetap bertahan: *Random Forest structural-only* menjadi model terbaik secara F1, sedangkan model *text-only* tidak cukup kompetitif. Dengan kata lain, sinyal kompromi yang berasal dari struktur halaman lebih stabil daripada sinyal linguistik pada tugas deteksi SEO-spam judi online.

Hasil tersebut dapat dijelaskan oleh karakteristik serangan. Banyak halaman spam di domain akademik tidak ditulis seperti artikel biasa, tetapi berupa hasil injeksi pada template yang sudah ada. Penyerang dapat mengubah kata kunci, menyisipkan variasi istilah, atau meminimalkan konten teks agar tidak mudah dikenali. Namun, pola seperti lonjakan external links, anchor yang mencurigakan, keberadaan *hidden div*, atau kepadatan elemen non-standar tetap meninggalkan jejak yang lebih konsisten. Temuan ini sejalan dengan literatur mutakhir yang menunjukkan pentingnya sinyal HTML, URL, dan struktur halaman, serta mengingatkan bahwa sebagian fitur dapat bersifat *dataset*-dependent bila tidak diuji pada skenario generalisasi yang lebih realistis [2], [6], [10].

Meski demikian, model *hybrid* tetap relevan. Ketika fitur teks digabungkan dengan fitur struktural, model memperoleh konteks tambahan yang berguna untuk kasus-kasus ambigu. *Hybrid SVM* dalam penelitian ini menawarkan *precision* tertinggi, artinya ketika model memprediksi suatu halaman sebagai *malicious*, prediksi tersebut cenderung benar. Nilai ini penting jika sistem dipakai untuk prioritisasi *triase insiden* atau penyaringan awal sebelum verifikasi manual. Dari sisi rekayasa sistem, model linear juga lebih mudah diaudit dibandingkan model *ensemble* yang lebih kompleks.

4.3 Perbandingan dengan evaluasi acak

Selama pengembangan, penelitian ini juga menemukan bahwa *random split* menghasilkan F1 yang sangat tinggi, mendekati 0,98, tetapi angka tersebut tidak dipertahankan sebagai klaim utama karena bersifat *inflated*. Hal ini terjadi karena halaman dari domain yang sama dapat muncul pada set latih dan uji, sehingga model pada dasarnya belajar mengenali template domain. Saat evaluasi diganti menjadi domain-holdout, F1 turun ke kisaran yang lebih realistis. Penurunan ini bukan kelemahan, melainkan bukti bahwa protokol evaluasi yang lebih ketat memberikan estimasi generalisasi yang lebih jujur. Dalam konteks studi web berbasis domain, temuan ini penting karena banyak karya praktis masih menggunakan *random split* meskipun data memiliki korelasi intra-domain yang kuat.

4.4 Analisis kesalahan

Analisis kesalahan pada *hybrid SVM* menunjukkan 25 prediksi salah pada test set, terdiri atas 7 *false positive* dan 18 *false* negative. Kasus *false positive* umumnya muncul pada halaman yang secara tekstual tampak normal, tetapi memiliki sinyal struktural mirip spam, misalnya jumlah link keluar yang tinggi atau anchor yang tampak tidak wajar. Sebaliknya, *false* negative banyak terjadi pada halaman judol yang muatan teksnya sangat minim, menggunakan istilah tersamar, atau hanya menampilkan jejak kompromi tipis pada struktur halaman.

Dari sudut pandang operasional, pola ini memberikan dua pelajaran. Pertama, jika tujuan utama adalah meminimalkan *false* alert, model dengan *precision* tertinggi seperti *hybrid SVM* masih layak dipilih sebagai model triase. Kedua, untuk mengejar *recall* yang lebih tinggi, pengembangan berikutnya dapat menambahkan fitur dari path URL, token meta tag, pola JavaScript, atau representasi visual halaman. Penelitian ini juga menunjukkan bahwa kesalahan tidak selalu berasal dari kelemahan model, tetapi dapat pula disebabkan oleh sifat konten spam yang sengaja dirancang agar menyerupai halaman normal.

Berikut adalah cuplikan kode HTML mentah yang diekstraksi langsung dari situs-situs yang salah diprediksi oleh model sebagai *false positive* dan *false negative*.

```
<!-- Kumpulan tautan keluar sah yang memicu deteksi link-farming (False Positive) -->
<div class="indexing-list">
  <a href="https://search.crossref.org/?q=10.15294" target="_blank" rel="noopener noreferrer"><img src="/public/site/images/crossref.png" alt="Crossref"></a>
  <a href="https://doaj.org/search/journals?source={"query":{"query_string":{"query":"unnes"%..."><img src="/public/site/images/doaj.png" alt="DOAJ"></a>
  <a href="https://sinta.kemdikbud.go.id/journals/?q=unnes" target="_blank"><img src="/public/site/images/sinta.png" alt="SINTA"></a>
  <a href="https://scholar.google.com/scholar?q=site:journal.unnes.ac.id" target="_blank">Google Scholar</a>
</div>
```

Gambar 1 Kode untuk *false positive*

 ISSN:2302-8149
e-ISSN:2540-9719

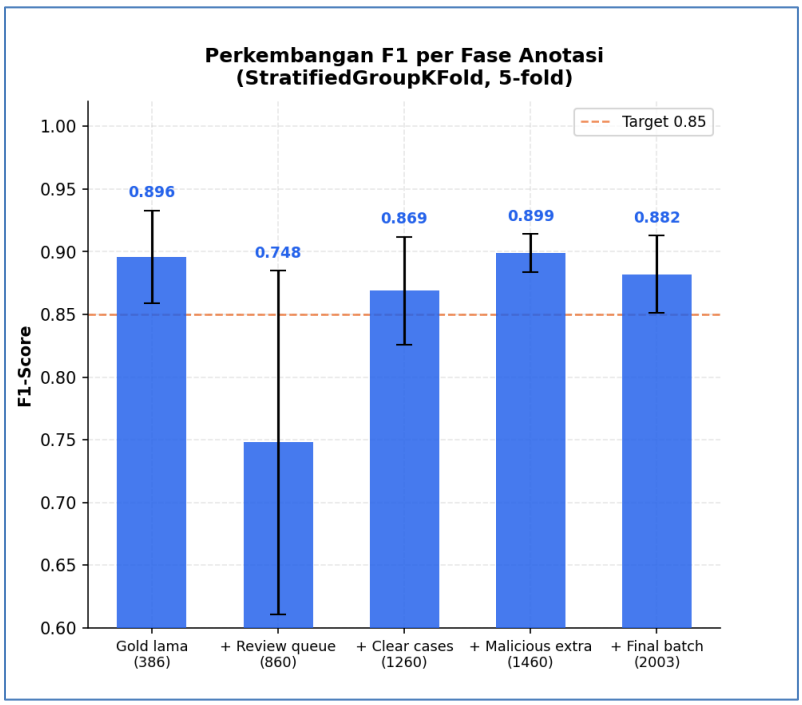
```
<!-- Injeksi Anchor text judi online di tengah-tengah OJS Footer (False Negative) -->
<div class="pkp_structure_footer_content">
    <p>Copyright &copy; 2026 LPPM Sekolah Tinggi Ilmu Ekonomi PGRI Sukabumi.</p>
    <!-- Injeksi tersembunyi dimulai -->
    <div class="seo-spam-injection">
        <a href="https://www.kingstribeca.com/">situs slot gacor</a>
        <a href="https://www.nabung77.com/">nabung77</a>
        <a href="https://thetopknotkitchen.com/">slot gacor</a>
    </div>
    <!-- Injeksi tersembunyi selesai -->
    <p>Journal is published under Open Access Policy.</p>
</div>
```

Gambar 2 Kode *false negative*

Pada Gambar 1 yang *false positive* Situs itu merupakan portal raksasa OJS. Karena model menggunakan fitur `*suspicious_links*` (kepadatan tautan keluar), ia menghukum halaman ini akibat banyaknya tautan sah menuju mesin pengindeks. Sedangkan pada Gambar 2 yang *false negative* penyerang menyisipkan deretan tautan (HTML Anchor) kasino ke dalam elemen *footer* jurnal. Walaupun ini adalah injeksi judol tulen, teks utamanya masih sangat sarat dengan kata-kata OJS ("pengabdian", "universitas").

4.5 Visualisasi hasil

Untuk memberikan gambaran yang lebih jelas mengenai dinamika performa model selama proses penyusunan dataset, evaluasi metrik kinerja terus dipantau pada setiap tahapan. Pergerakan nilai rerata F1-score yang merepresentasikan fluktuasi dan peningkatan performa klasifikasi pada masing-masing fase kurasi data dapat diamati secara langsung melalui visualisasi pada Gambar 3 berikut.

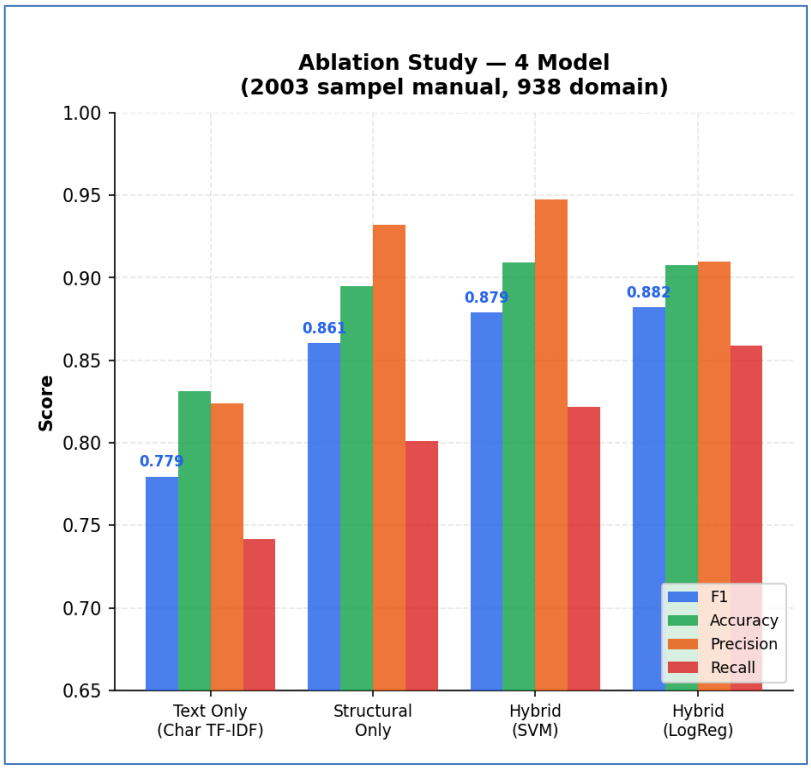


Gambar 3 Perkembangan F1 per fase anotasi

 ISSN:2302-8149
e-ISSN:2540-9719

Visualisasi pada Gambar 3 menunjukkan perkembangan rerata *F1-score* selama beberapa fase anotasi *dataset* menggunakan evaluasi *StratifiedGroupKFold* berbasis domain. Nilai F1 sempat turun pada fase "*Review queue*" menjadi 0,748 akibat bertambahnya sampel yang lebih beragam dan sulit diklasifikasikan. Setelah proses kurasi dan pembersihan *dataset*, performa kembali meningkat hingga mencapai 0,899 pada fase "*Malicious extra*". Pada *dataset* final, model memperoleh F1 sebesar 0,882 dan tetap berada di atas target evaluasi 0,85, menunjukkan bahwa *dataset* akhir memiliki kualitas anotasi yang lebih stabil dan realistis.

Selain memantau perkembangan metrik selama fase anotasi, analisis lanjutan juga dilakukan untuk mengevaluasi efektivitas masing-masing kelompok fitur dan mengukur signifikansi peningkatan performa dari model awal hingga model final. Komparasi performa melalui ablation study, evaluasi tiga generasi pengembangan model, beserta rincian proporsi distribusi label kelas pada dataset akhir dapat dilihat secara komprehensif pada visualisasi di Gambar 4 berikut.

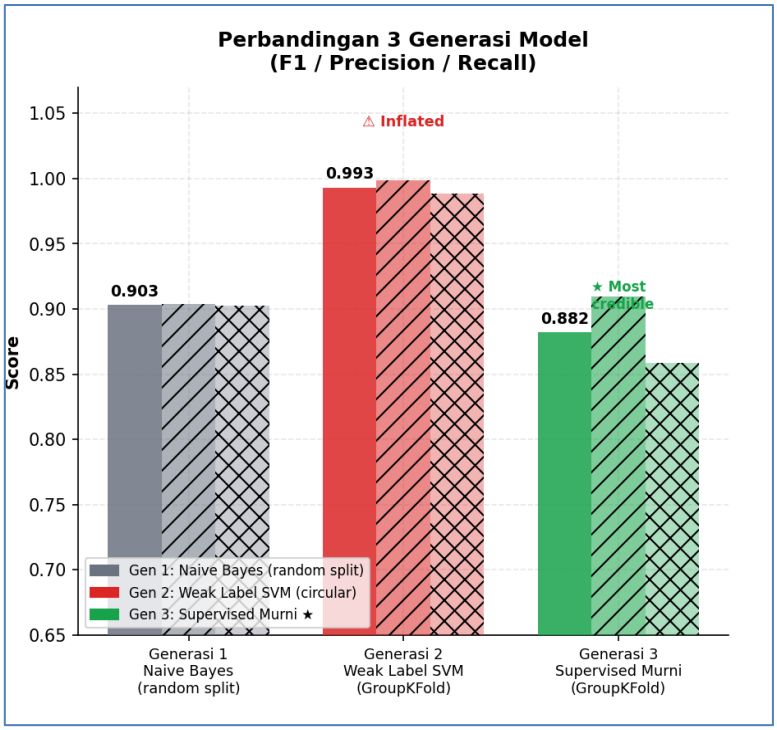


Gambar 4 *Ablation* study

Visualisasi pada Gambar 4 memperlihatkan hasil *ablation study* terhadap empat pendekatan model. Pendekatan *structural-only* menghasilkan performa terbaik dengan F1 sekitar 0,884, menunjukkan bahwa fitur struktural lebih efektif dibanding fitur tekstual dalam mendeteksi SEO-spam judi. Sementara itu, *Hybrid SVM* menghasilkan *precision* tertinggi, menandakan bahwa kombinasi fitur tekstual dan struktural mampu mengurangi *false positive* pada proses klasifikasi.

Untuk mengevaluasi sejauh mana peningkatan dan stabilitas performa yang dihasilkan dari perbaikan metode evaluasi serta penyempurnaan pelabelan data, perbandingan kinerja dari tiga generasi pengembangan model dapat dilihat secara rinci pada Gambar 5 berikut.

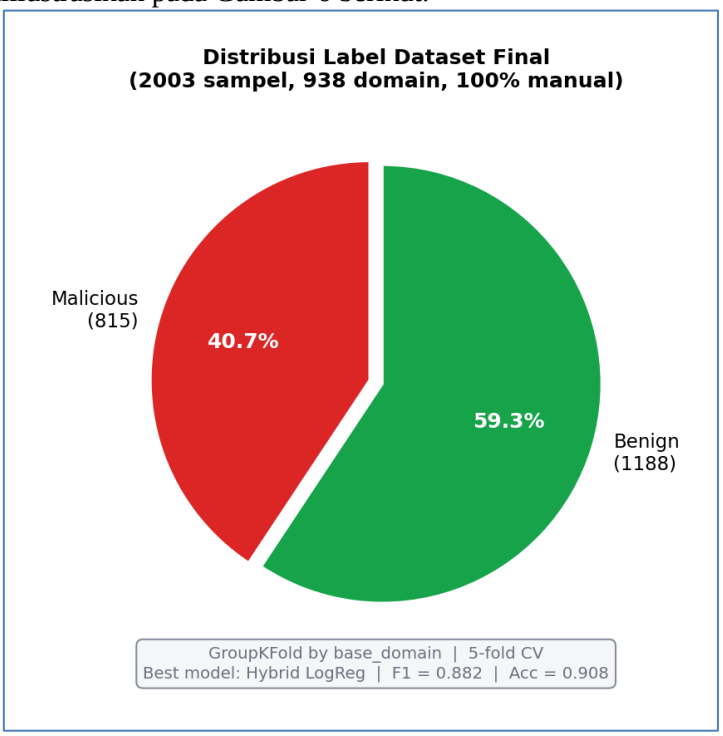
 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 5 Perbandingan 3 generasi model

Visualisasi pada Gambar 5 menunjukkan perbandingan tiga generasi model selama pengembangan penelitian. Generasi awal dengan *random split* dan weak-label menghasilkan F1 yang sangat tinggi hingga 0,993, tetapi performa tersebut bersifat *inflated* akibat potensi data *leakage* antar-domain. Setelah menggunakan GroupKFold berbasis domain dan anotasi manual penuh, performa menjadi lebih realistis dengan F1 sebesar 0,882 pada generasi ketiga.

Sebagai gambaran mengenai komposisi kelas data yang digunakan dalam tahap pelatihan dan pengujian akhir, proporsi antara halaman bersinyal *malicious* dan halaman *benign* pada dataset yang telah dianotasi penuh diilustrasikan pada Gambar 6 berikut.



Gambar 6 Distribusi label *dataset* final

 ISSN:2302-8149
e-ISSN:2540-9719

Visualisasi pada Gambar 6 menunjukkan distribusi label pada *dataset* final yang terdiri atas 40,7% sampel *malicious* dan 59,3% sampel benign. Distribusi ini menunjukkan kondisi moderately imbalanced namun masih representatif terhadap kondisi nyata pada domain akademik, di mana halaman benign lebih dominan dibanding halaman SEO-spam judi.

4.6 Keterbatasan penelitian

Penelitian ini memiliki beberapa keterbatasan. Pertama, meskipun seluruh label pada *dataset* final dibuat secara manual, proses anotasi masih melibatkan satu anotator sehingga potensi subjektivitas tetap ada. Kedua, *dataset* difokuskan pada domain akademik Indonesia, sehingga generalisasi ke domain pemerintah, komersial, atau media belum bisa diasumsikan langsung. Ketiga, fitur yang digunakan masih didominasi ringkasan teks dan struktur HTML dasar; fitur yang lebih kaya seperti graph connectivity, informasi jaringan, atau representasi visual halaman belum dieksplorasi.

Meski demikian, penelitian ini tetap memberikan kontribusi yang kuat karena memisahkan data final dari arsip *weak* label, mendokumentasikan protokol evaluasi berbasis domain, serta menunjukkan trade-off yang jelas antara model dengan F1 tertinggi dan model dengan *precision* tertinggi. Dengan demikian, hasil penelitian ini dapat dijadikan fondasi untuk benchmark lanjutan maupun sistem prioritisasi insiden yang lebih matang.

5 Kesimpulan

Penelitian ini mengembangkan sistem deteksi SEO-spam judi pada domain akademik Indonesia dengan memanfaatkan kombinasi fitur tekstual dan struktural serta protokol evaluasi domain-holdout yang ketat. *Dataset* final berisi 2.003 halaman dari 938 domain .ac.id dan seluruh label pada *dataset* akhir ditetapkan secara manual. Hasil evaluasi menunjukkan bahwa *Random Forest* berbasis fitur struktural mencapai F1 terbaik sebesar 0,8844, sedangkan *hybrid SVM* mencapai *precision* tertinggi sebesar 0,9205 dengan F1 0,8663.

Referensi

[1] L. Tang and Q. H. Mahmoud, "*A Survey of Machine Learning-based Solutions for Phishing Website Detection*," Machine Learning and Knowledge *Extra*ction, Vol. 3, No. 3, pp. 672-694, 2021, DOI: 10.3390/make3030034.

[2] Y. Tian, Y. Yu, J. Sun, and Y. Wang, "*From Past to Present: A Survey of Malicious URL Detection Techniques, Datasets and Code Repositories*," Computer Science Review, Vol. 58, p. 100810, 2025, DOI: 10.1016/j.cosrev.2025.100810.

[3] M. Sánchez-Paniagua, E. Fidalgo, E. Alegre, and R. Alaiz-Rodríguez, "*Phishing Websites Detection using a Novel Multipurpose Dataset and Web Technologies Features*," Expert Systems with Applications, Vol. 207, p. 118010, 2022, DOI: 10.1016/j.eswa.2022.118010.

[4] A. K. Jha, R. Muthalagu, and P. M. Pawar, "*Intelligent Phishing Website Detection using Machine Learning*," Multimedia Tools and Applications, Vol. 82, pp. 29431-29456, 2023, DOI: 10.1007/s11042-023-14731-4.

[5] C. Opara, Y. Chen, and B. Wei, "*Look Before You Leap: Detecting Phishing Web Pages by Exploiting Raw URL and HTML Characteristics*," Expert Systems with Applications, Vol. 236, p. 121183, 2024, DOI: 10.1016/j.eswa.2023.121183.

[6] S. Sheikhi and P. Kostakos, "*Safeguarding Cyberspace: Enhancing Malicious Website Detection with PSO Optimized XGBoost and Firefly-based Feature Selection*," Computers & Security, Vol. 142, p. 103885, 2024, DOI: 10.1016/j.cose.2024.103885.

[7] W. Li, S. Manickam, Y.-W. Chong, W. Leng, and P. Nanda, "*A State-of-the-Art Review on Phishing Website Detection Techniques*," IEEE Access, Vol. 12, pp. 187976-188012, 2024, DOI: 10.1109/ACCESS.2024.3514972.

[8] S. Kavya and D. Sumathi, "*Staying Ahead of Phishers: A Review of Recent Advances and Emerging Methodologies in Phishing Detection*," Artificial Intelligence Review, Vol. 58, art. No. 50, 2025, DOI: 10.1007/s10462-024-11055-z.

[9] G. S. Collins, K. G. M. Moons, P. Dhiman, R. D. Riley, A. L. Beam, B. Van Calster, et al., "*TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models that Use*

 ISSN:2302-8149
e-ISSN:2540-9719

Regression or Machine Learning Methods," BMJ, Vol. 385, p. e078378, 2024, DOI: 10.1136/bmj-2023-078378.

[10] M. Mia, D. Derakhshan, and M. M. A. Pritom, "*Can Features for Phishing URL Detection Be Trusted Across Diverse Datasets? A Case Study with Explainable AI*," in Proceedings of the 2nd International Conference on AI Foundation Models and Software Engineering, 2025, DOI: 10.1145/3704522.3704532.

[11] B. Wang, "*Malicious URL Detection with Explainable Machine Learning Techniques*," in Proceedings of the 2nd International Conference on Informatics, Education and Computer Technology Applications (IECA '25), pp. 293-299, 2025, DOI: 10.1145/3732801.3732854.

[12] J. L. Wilk-Jakubowski, Ł. Pawlik, G. Wilk-Jakubowski, and A. Sikora, "*Machine Learning and Neural Networks for Phishing Detection: A Systematic Review (2017–2024)*," Electronics, Vol. 14, No. 18, p. 3744, 2025, DOI: 10.3390/electronics14183744.

[13] R. Liu, Y. Wang, Z. Guo, H. Xu, Z. Qin, W. Ma, and F. Zhang, "*TransURL: Improving Malicious URL Detection with Multi-Layer Transformer Encoding and Multi-Scale Pyramid Features*," Computer Networks, Vol. 253, p. 110707, 2024, DOI: 10.1016/j.comnet.2024.110707.

[14] R. Liu, Y. Wang, H. Xu, Z. Qin, F. Zhang, Y. Liu, and Z. Cao, "*PMANet: Malicious URL Detection via Post-Trained Language Model Guided Multi-Level Feature Attention Network*," Information Fusion, 2025, DOI: 10.1016/j.inffus.2024.102638.

[15] A. E. Omolara and M. Alawida, "*DaE2: Unmasking Malicious URLs by Leveraging Diverse and Efficient Ensemble Machine Learning for Online Security*," Computers & Security, Vol. 148, p. 104170, 2025, DOI: 10.1016/j.cose.2024.10417.