



Pengembangan Web Scraper Menggunakan Algoritma Dfs untuk Analisis Penyusupan Situs Judi Online

Angga Rizki Ramadhani^{*1}, Moh. Dasuki², Lutfi Ali Muharom³

Email: ¹skripsiangga50@gmail.com, ²moh.dasuki22@unmuhjember.ac.id, ³lutfi.muharom@unmuhjember.ac.id

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Diterima: 12 Januari 2026 | Direvisi: 21 Februari 2026 | Disetujui: 27 April 2026

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Penyusupan *judi online* ke dalam domain resmi seperti ac.id, go.id, my.id, .id, dan sch.id menjadi permasalahan serius karena dapat merusak reputasi institusi, mengganggu keamanan informasi, dan menurunkan kepercayaan publik. Salah satu tujuan umum dari praktik penyusupan ini adalah untuk meningkatkan peringkat SEO situs judi tersebut dengan memanfaatkan otoritas domain resmi. Penelitian ini bertujuan untuk mengembangkan sebuah *web scraper* menggunakan algoritma *Depth First Search* DFS yang mampu melakukan penelusuran mendalam pada struktur website serta menganalisis penyusupan situs judi online secara otomatis. Sistem yang dibangun melakukan proses *crawling*, ekstraksi elemen html yang mengandung *link judi online*, analisis teks, dan pencocokan pola kata kunci. Pengujian dilakukan melalui beberapa tahap, dimulai dari uji awal pada 40 domain hingga pengujian akurasi formal pada 50 domain. Hasil menunjukkan bahwa sistem mampu bekerja stabil dan konsisten dalam melakukan penelusuran menggunakan DFS. Tingkat akurasi yang diperoleh mencapai 98%, dengan 49 hasil analisis sesuai dan 1 kesalahan deteksi. Evaluasi lanjutan pada 60 domain tambahan menghasilkan peningkatan akurasi menjadi 98,33% setelah dilakukan penyempurnaan pada modul analisis.

Kata kunci: *Web Scraper, DFS, Crawling, Analisis Penyusupan, Judi Online.*

Web scraper development using dfs algorithm for online gambling site intrusion analysis

Abstract

The infiltration of online gambling into official domains such as ac.id, go.id, my.id, .id, and sch.id has become a serious issue because it can damage institutional reputation, disrupt information security, and reduce public trust. One of the common objectives of this infiltration practice is to boost the SEO ranking of the gambling sites by exploiting the authority of official domains. This research aims to develop a web scraper using the Depth First Search DFS algorithm capable of performing deep traversal on website structures and automatically analyzing the infiltration of online gambling links. The system conducts crawling, extraction of html elements containing gambling-related links, text analysis, and keyword pattern matching. Testing was carried out in several stages, starting with an initial evaluation on 40 domains followed by a formal accuracy test on 50 domains. The results show that the system performs stably and consistently when conducting traversal using DFS. The achieved accuracy reached 98%, with 49 correct analyses and 1 detection error. Further evaluation on an additional 60 domains increased the accuracy to 98,33% after improvements were applied to the analysis module.

Keywords: *Web Scraper, DFS, Crawling, Intrusion Analysis, Online Gambling.*

1. PENDAHULUAN

Perjudian telah menjadi salah satu aktivitas kontroversial yang berkembang pesat di era modern. Meskipun sering dikaitkan dengan hiburan, aktivitas ini memiliki berbagai dampak negatif yang signifikan, termasuk pada aspek ekonomi, sosial, dan psikologis. Dalam beberapa tahun terakhir, perkembangan teknologi digital telah mempermudah akses ke aktivitas perjudian, terutama melalui platform daring. Fenomena ini semakin memprihatinkan, terutama karena banyak situs perjudian *online* yang menyusup ke situs web resmi, seperti milik sekolah, kampus, atau lembaga pemerintahan [1].

Penyusupan *link* atau konten perjudian ke situs-situs resmi menjadi masalah yang sangat serius. Situs resmi yang biasanya digunakan sebagai sumber informasi terpercaya, ketika terkontaminasi oleh *link* perjudian, dapat kehilangan kredibilitas dan kepercayaan dari pengunjungnya. bahkan, penyusupan ini sering kali tidak terdeteksi oleh pengelola situs, mengingat teknik yang digunakan oleh pelaku sangat canggih, seperti melalui injeksi kode atau penggunaan tautan tersembunyi *hidden links*, keberadaan *link* perjudian di situs resmi juga dapat memicu berbagai masalah lainnya. Selain merusak reputasi, hal ini berpotensi melanggar undang-undang, terutama di negara-negara yang melarang aktivitas perjudian [2].

Salah satu alasan utama penyebaran *link* perjudian adalah strategi pemasaran agresif dari operator perjudian untuk meningkatkan seo, atau menghindari pemblokiran langsung [3]. Dengan tingginya tingkat kunjungan pada situs resmi, seperti situs pendidikan atau pemerintahan, peluang pengguna mengakses *link* perjudian menjadi lebih besar. Hal ini memberikan keuntungan bagi operator *judi*, namun berdampak buruk bagi pemilik situs resmi. Selain itu, keberadaan *link* perjudian di situs resmi dapat membahayakan pengunjung situs tersebut. *Link* tersebut sering kali digunakan untuk memancing korban dalam berbagai skema penipuan atau serangan siber, seperti phishing atau penyebaran *malware* [4].

Berdasarkan data dari komdigi, sebanyak 683 website pemerintah dan lembaga pendidikan dilaporkan telah disusupi konten *judi online* dalam rentang waktu 1 Januari 2022 hingga 13 Februari 2023. Dari jumlah tersebut, 461 adalah situs pemerintahan dengan domain go.id, sementara 222 lainnya adalah situs pendidikan dengan domain ac.id. komdigi mengklaim telah menangani permasalahan ini melalui teknik *web crawling* serta laporan dari masyarakat [5].

Teknologi seperti *web scraper* yang mampu melakukan perayapan otomatis untuk mengidentifikasi *link* mencurigakan sering kali belum diterapkan secara luas [6]. Tidak hanya sekadar *web scraper*, alat yang dirancang dalam penelitian ini juga menggunakan beberapa algoritma atau metode pengumpulan data seperti dfs, *web crawling*, akurasi klasifikasi biner sehingga menjadi solusi yang lebih *powerful*.

2. METODE PENELITIAN

Sebelum proses analisis dapat dilakukan, diperlukan fakta atau pengetahuan yang menjelaskan bagaimana konten perjudian dapat menyusup ke dalam situs web secara tersembunyi. informasi semacam ini tidak selalu terlihat secara langsung, karena penyusupan sering kali dilakukan melalui tautan atau elemen halaman yang tampak normal. untuk memperoleh pengetahuan tersebut, dibutuhkan usaha pengumpulan data secara terstruktur dari halaman situs web yang memiliki keterkaitan. *web scraper* digunakan sebagai alat bantu untuk mengekstraksi elemen-elemen yang dicurigai mengandung konten perjudian, agar informasi yang sebelumnya tidak terlihat dapat ditelusuri dan dianalisis secara menyeluruh.

2.1. Pengumpulan Data

Tahapan ini menjelaskan sumber dan bentuk data yang digunakan dalam pengumpulan data. data diperoleh melalui proses penelusuran terhadap sejumlah halaman situs web yang menjadi target. Informasi yang dikumpulkan meliputi. Url halaman sumber, isi teks halaman, tautan keluar, kata kunci terkait perjudian [7].

Web dengan domain .ac.id dan .go.id merupakan jenis situs yang cukup rentan menjadi sasaran penyusupan konten perjudian *online*. penyusupan tersebut biasanya ditandai dengan kemunculan kata toto88, ayo deposit, pada pencarian google, yang disisipkan secara tersembunyi dalam halaman situs resmi. komdigi mencatat bahwa sebanyak 5.128.871 konten perjudian telah berhasil ditangani dalam kurun waktu 2017 hingga November 2024, yang menunjukkan bahwa ancaman ini masih terus berlangsung secara masif [8], temuan tersebut menjadi salah satu pertimbangan dalam pengumpulan data.

Sumber data diperoleh dari sejumlah situs web publik yang menunjukkan ciri-ciri memiliki indikasi penyusupan judi online. untuk membantu menemukan situs-situs yang berpotensi mengandung judi online tersebut, digunakan pendekatan *Open Source Intelligence* dengan menerapkan teknik pencarian terbuka berbasis pola tertentu melalui platform seperti google, censys, shodan, dan fofa yang umum digunakan dalam aktivitas pemantauan siber [9], web yang ditemukan melalui metode ini selanjutnya menjadi objek analisis untuk melihat adanya keterkaitan dengan konten perjudian tersembunyi.

2.2. Perancangan Perangkat Lunak

Perancangan *web scraper* ini dirancang untuk memastikan sistem dapat berfungsi secara optimal dalam mendeteksi dan menganalisis keberadaan situs *judi online* yang menyusup pada domain tertentu. Sistem ini dirancang dengan alur kerja yang terstruktur, mulai dari proses *input* url, penelusuran halaman menggunakan teknik *crawling* atau dfs, hingga menghasilkan keluaran berupa daftar *link* atau tag html yang mencirikan situs *judi online*.

1. Analisis

Sistem yang akan di buat bertujuan untuk mendeteksi penyusupan situs *judi online* pada domain yang sudah di pilih. Sistem ini ditujukan untuk pengguna yang bertugas menganalisis hasil *scraper* data agar proses pendeteksian lebih efektif. Analisis kebutuhan sistem ini terbagi menjadi dua, yaitu analisis kebutuhan fungsional dan analisis kebutuhan non fungsional.

Pada analisis desain sistem, dilakukan perancangan teknis berdasarkan kebutuhan yang telah diidentifikasi untuk memastikan sistem dapat berfungsi sesuai dengan tujuan. Analisis desain mencakup dua aspek utama, yaitu desain antarmuka pengguna dan alur proses sistem.

Penggunaan domain resmi sering kali disalahgunakan oleh pihak tidak bertanggung jawab untuk menyisipkan link *judi online*. Masalah ini menyebabkan kerugian, baik secara moral maupun reputasi, bagi pemilik domain resmi. Selain itu, pendeteksian situs *judi online* pada domain domain ini secara manual memakan waktu, tidak efisien, dan rentan terhadap kesalahan.

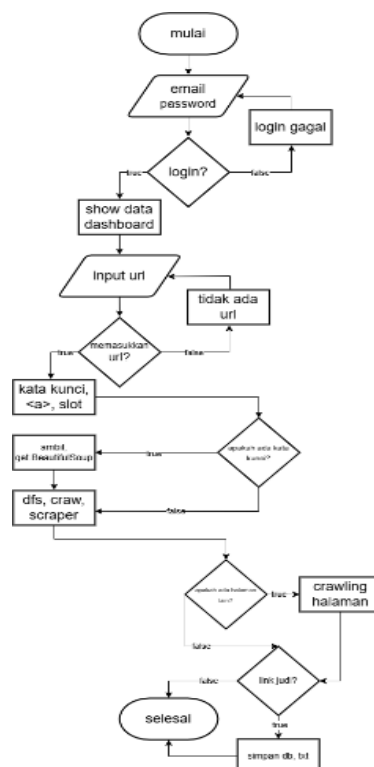
2. Desain

Antarmuka ini dirancang untuk memungkinkan pengguna melakukan proses *web scraper* secara terarah dan sesuai. Pada bagian ini, terdapat sebuah kolom input yang digunakan untuk memasukkan url domain yang akan menjadi target *scraping*. Setelah url dimasukkan, pengguna dapat menekan tombol *Shot Scraper Web* untuk memulai proses *scraping*. Sistem akan melakukan *crawling* menggunakan algoritma *depth first search* pada domain tersebut untuk mencari *link* atau teks yang relevan dengan kata kunci tertentu, seperti istilah yang sering muncul pada situs slot atau *judi online*. Dapat dilihat pada Gambar 1.



Gambar 1. Wireframe Web Scraper

Sistem dimulai dengan pengguna yang diminta untuk memasukkan url yang akan dianalisis. Jika pengguna berhasil memasukkan url, maka sistem akan melanjutkan proses dengan meminta daftar kata kunci yang benar, yang dalam hal ini adalah kata kunci yang berhubungan dengan situs *judi online*. Setelah itu, fungsi *cari_judi* url akan dijalankan untuk memulai proses pencarian pada url yang diberikan. Dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Web Scraper

Sistem memiliki beberapa fungsi utama yang dapat diakses oleh aktor human setelah melakukan proses *login*. Fungsi fungsi ini meliputi buat *login*, data domain, *web scraper*, data target, data analisis, dan tabel human. Hubungan perluasan digunakan untuk menunjukkan fitur opsional yang memperluas perilaku *use case* utama, seperti fitur perbarui domain, hapus domain, mencari *scraper*, dan hapus human, yang hanya dijalankan dalam kondisi tertentu. Sementara itu, hubungan inklusi menggambarkan bahwa beberapa fungsi, seperti verifikasi link dan baca file, merupakan bagian penting dari proses tertentu, misalnya pada *web scraper* dan data target. Dapat dilihat pada Gambar 3.



Gambar 3. Use Case Sistem Web Scraper

3. Develop

Struktur kode dalam sistem ini dirancang dengan memperhatikan modularitas dan pemisahan tanggung jawab setiap komponen untuk meningkatkan keterbacaan, pemeliharaan, dan pengembangan lebih lanjut.

4. Testing

Black box testing adalah metode pengujian yang berfokus pada fungsi sistem tanpa melihat struktur internal. Pada sistem ini, pengujian dilakukan untuk memastikan proses *login*, *input* domain, *scraper* data, dan analisis berjalan sesuai harapan. Hasil pengujian menunjukkan apakah sistem merespons *input* dengan benar dan menghasilkan *output* sesuai kebutuhan pengguna [10].

2.3. Implementasi

Pola yang mencirikan situs *judi online* dapat dilihat dari elemen-elemen tertentu yang muncul secara konsisten dalam struktur html halaman web. Beberapa ciri khas yang dapat digunakan untuk mengidentifikasi situs judi adalah adanya kata-kata atau frasa tertentu yang terkait dengan perjudian, seperti “slot”, “taruhan”, “poker”, “casino”, dan “*judi online*” [11], Selain itu, tag html yang mengandung atribut atau isi yang mengarah ke permainan judi juga menjadi indikator penting, pada situs-situs yang terindikasi menyusupkan konten *judi*, ditemukan pula penggunaan gambar, iklan [12], atau elemen visual lainnya yang mengarah ke situs atau platform *judi online*. Dapat dilihat pada Gambar 4.



Gambar 4. Pola Pencarian Google Dorking

Pengujian akan dilakukan untuk memastikan bahwa sistem *scraper* yang dikembangkan dapat mendeteksi keberadaan indikasi situs *judi online* pada website target [13]. Salah satu situs yang akan digunakan dalam pengujian ini adalah situs resmi universitas muhammadiyah jember

Rencana Pengujian:

1. Target domain manajemen.unmuhjember.ac.id/index.php/id/
2. Searching for Target Domains: manajemen.unmuhjember.ac.id
3. Jika data slot di temukan maka *Scraper Finds Negative Data Things*: <a>

Untuk menguji akurasi sistem, dilakukan pengujian pada sejumlah domain dengan ekstensi tertentu, dengan total 50 situs yang diuji, masing-masing kategori domain terdiri dari 10 situs. Dapat dilihat pada Tabel 1.

Tabel 1. Jumlah Data per Domain

ac.id, go.id, my.id, .id, dan sch.id	
Domain	Jumlah
ac.id	10
go.id	10
my.id	10
.id	10
sch.id	10

Tingkat akurasi dari sistem ini dihitung menggunakan rumus 1 berikut:

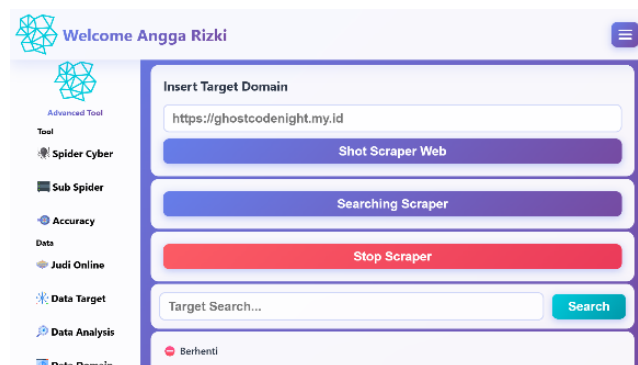
$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

Evaluasi pada tahap ini difokuskan untuk menilai hasil akurasi sistem *web scraper* terhadap data uji, data yang digunakan berasal dari hasil *scraper* terhadap sejumlah domain target. Dalam evaluasi ini, sistem akan diuji berdasarkan ketepatan hasil deteksi, ketepatan domain, jumlah kesalahan deteksi.

3. HASIL DAN PEMBAHASAN

3.1. Gambaran Umum Sistem

Website yang dihasilkan memiliki beberapa fitur utama, seperti menjalankan proses *scraper*, menampilkan hasil analisis, menampilkan data akurasi, serta mengelola data domain dan pengguna. seluruh fitur ini ditampilkan dalam antarmuka terintegrasi yang dapat diakses melalui menu utama aplikasi. Dapat dilihat pada Gambar 5.



Gambar 5. Dashboard Web Scraper

Sistem *web scraper* menjalankan proses mulai dari autentikasi pengguna hingga penyajian hasil analisis. setelah pengguna berhasil masuk, sistem menerima *input* domain yang akan diperiksa dan langsung memulai proses penelusuran halaman. selama proses tersebut, sistem mengunjungi setiap halaman yang ditemukan, mengambil elemen teks serta tautan keluar, dan mendeteksinya berdasarkan indikator penyusupan *judi online*.

3.2. Hasil Pengujian Sistem

Pengujian pada tahap ini mengacu pada metode *black box testing*, pengujian dilakukan untuk memastikan bahwa setiap halaman dan fitur utama pada sistem berfungsi sesuai dengan kebutuhan. Dapat dilihat pada Tabel 2.

Tabel 2. Black Box Testing

No	Halaman / Fitur	Skenario Pengujian	Hasil Diharapkan	Hasil Pengujian
1	Login	Pengguna memasukkan <i>username</i> dan <i>password</i>	Sistem menampilkan halaman utama	Sesuai
2	Dashboard Scraper	Pengguna memasukkan domain target dan menekan tombol <i>Shot scraper web</i>	Sistem melakukan proses <i>scraper</i> dan menampilkan hasil	Sesuai
3	Halaman Judi Online	Memilih <i>dropdown</i>	Menampilkan hasil <i>judi online</i>	Sesuai
4	Data Target	Menampilkan daftar hasil <i>scraper</i> domain	Data tampil sesuai hasil <i>scraper</i>	Sesuai
5	Data Analysis	Menampilkan hasil analisis domain <i>judi online</i> dan domain asli	Sistem menampilkan daftar domain dan hasil analisis	Sesuai

6	Data Domain	Menampilkan domain, click <i>delete, update</i>	Sistem menampilkan domain, bisa <i>delete, maupun update</i> data	Sesuai
7	Table Human	Menampilkan data pengguna, dapat menghapus pengguna	Sistem dapat menampilkan dan menghapus data dengan benar	Sesuai
8	Create Human	Dapat menambah pengguna baru	Menyimpan data dengan benar	Sesuai
9	Log Out	Pengguna click	Keluar Halaman	Sesuai
10	Sidebar	Pengguna click	Buka tutup <i>sidebar</i>	Sesuai

Waktu eksekusi DFS berbeda pada setiap domain. domain *go.id* memiliki waktu *crawling* paling lama yaitu 30 menit, sedangkan *my.id* menjadi yang paling cepat dengan waktu 16 menit. perbedaan ini muncul karena variasi jumlah halaman internal yang berhasil diakses pada masing-masing situs. semakin banyak tautan yang ditemukan, semakin lama proses *crawling* diselesaikan. Dapat dilihat pada Tabel 3.

Tabel 3. Waktu Eksekusi DFS

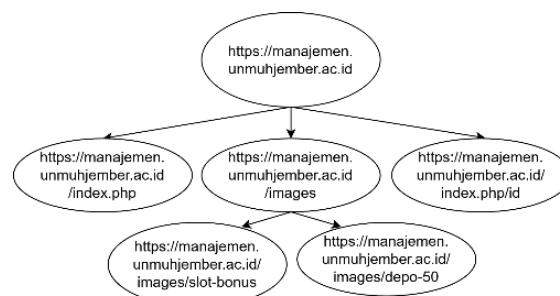
Domain	Waktu Eksekusi
ac.id	17 Menit
go.id	30 Menit
my.id	16 Menit
.id	18 Menit
sch.id	17 Menit
Total jumlah 50 domain	

Hasil pengujian menunjukkan bahwa setiap domain memiliki kedalaman penelusuran antara 3-4 level. nilai ini menggambarkan bahwa DFS yang diimplementasikan mampu menjangkau struktur halaman hingga level terdalam yang tersedia. kedalaman tertinggi ditemukan pada domain *ac.id, go.id, dan sch.id* dengan *depth* 4, sementara *my.id dan .id* memiliki *depth* 3. variasi ini secara langsung berpengaruh terhadap jumlah data yang berhasil dihimpun oleh sistem, karena semakin dalam struktur suatu domain, semakin banyak pula halaman berpotensi terlihat oleh *crawling*. Dapat dilihat pada Tabel 4.

Tabel 4. *traversal depth*

Domain	Kedalaman
ac.id	4
go.id	4
my.id	3
.id	3
sch.id	4
Total jumlah 50 domain	

Proses *crawling* yang dilakukan sistem pada setiap domain menghasilkan alur penelusuran yang dapat diamati secara langsung melalui struktur halaman yang berhasil dikunjungi. berdasarkan hasil pengujian, pola penelusuran yang terbentuk pada beberapa domain menunjukkan bagaimana sistem mengikuti tautan-tautan internal hingga kedalaman tertentu. representasi salah satu hasil *crawling* ditunjukkan pada Gambar 6.



Gambar 6. Alur DFS Pada Sebuah Website

Pengujian proses *crawling* dilakukan pada lima domain target yang memiliki karakteristik struktur halaman berbeda. pada pengujian ini, sistem mencatat jumlah halaman yang berhasil dikunjungi serta jumlah tautan yang ditemukan selama proses penelusuran. data tersebut kemudian dirangkum untuk melihat seberapa besar cakupan *crawling* pada masing-masing domain. hasil lengkap pengujian ditampilkan pada Tabel 5 berikut.

Tabel 5. Hasil Analisis DFS

Domain	Jumlah Halaman (V)	Jumlah Tautan (E)	Kompleksitas $O(V + E)$
ac.id	2.140	2.920	$O(5.060)$
go.id	4.000	6.500	$O(10.500)$
my.id	1.999	2.500	$O(4.499)$
.id	2.210	2.200	$O(4.410)$
sch.id	2.110	2.550	$O(4.660)$

Pada studi kasus pertama, sistem mendeteksi adanya indikasi kuat penyusupan *judi online* berdasarkan hasil analisis *scraper*. domain ini merupakan contoh situs dengan struktur halaman yang telah dimodifikasi atau disusupi oleh pihak tidak bertanggung jawab.

Cuplikan real hasil deteksi:

Searching for Target Domains: <https://bnn.go.id>

Scraper Finds Negative Data Things: `https://mudikgratis.dishub.jatimprov.go.id/-/slot-demo/`

Scraper Finds Negative Data Things: `Main Gacor`

Searching for Target Domains: <https://bnn.go.id/penulis/humas>

Pada studi kasus kedua, sistem memberikan hasil analisis terhadap domain yang tidak menunjukkan adanya tanda-tanda penyusupan *judi online*. studi ini dilakukan untuk memastikan bahwa mekanisme deteksi tidak hanya mampu menemukan indikasi positif, tetapi juga dapat mengenali kondisi situs yang benar-benar bersih.

Cuplikan real hasil deteksi:

Searching for Target Domains: <https://adpim.kalbarprov.go.id/welcome>

Searching for Target Domains: <https://adpim.kalbarprov.go.id/beranda>

Searching for Target Domains: <https://adpim.kalbarprov.go.id/beranda#profil>

Searching for Target Domains: <https://adpim.kalbarprov.go.id/sambutan-kepala-biro-administrasi-pimpinan.html>

Searching for Target Domains: <https://adpim.kalbarprov.go.id/sambutan-kepala-biro-administrasi-pimpinan.html#profil>

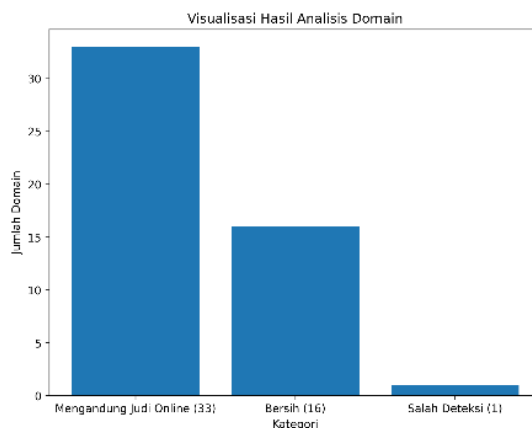
3.3. Hasil Pengumpulan Data

Hasil pengumpulan data merupakan keluaran dari proses penelusuran *scraper* yang di mana sistem *web scraper* menelusuri sejumlah halaman situs web dengan domain target untuk mencari domain yang berpotensi mengandung indikasi penyusupan *judi online*, seluruh hasil pengumpulan disusun dan direkapitulasi dalam Tabel 6.

Tabel 6. Rekap Hasil Pengumpulan Data

Jenis Domain	Jumlah Domain	Isi Teks Halaman Terkait Judi Online	Jumlah Domain Judi Online	Jumlah Kata Kunci Judi Online
ac.id	10	30.543	3.950	50.244
go.id	10	90.343	16.200	219.960
my.id	10	20.232	2.343	37.180
.id	10	3.405	1.975	4.200
sch.id	10	24.768	14.280	104.865

Setelah memaparkan domain uji beserta jenis temuan yang muncul pada masing-masing kategori, diperlukan pula penyajian visual untuk memberikan gambaran pada Gambar 7 yang lebih jelas mengenai distribusi hasil deteksi pada seluruh data uji. visualisasi ini tidak hanya menampilkan jumlah domain yang mengandung indikasi penyusupan maupun yang bersih, tetapi juga memperlihatkan kesalahan deteksi yang terjadi selama proses analisis.



Gambar 7. Visualisasi Hasil Analisis Domain

Disajikan hasil analisis temuan berdasarkan tiga pola penyusupan yang telah didefinisikan, analisis dilakukan untuk mengetahui seberapa sering pola-pola tersebut muncul pada hasil scraper, maupun di saat pencarian sumber data pada setiap kelompok domain, disajikan pada Tabel 7.

Tabel 7. Analisis Pola

Jenis Domain	Jumlah Domain	Dorking	Deface	Path
ac.id	10	4	4	80
go.id	10	4	6	1
my.id	10	3	1	0
.id	10	2	1	0
sch.id	10	4	4	1
Total	50	17	16	82

Sumber data pada penelitian ini diperoleh dari hasil pencarian terbuka dan melalui empat platform utama, keempat platform tersebut digunakan untuk menemukan domain yang menunjukkan potensi penyusupan *judi online*.

Pencarian dilakukan dengan menerapkan pola kueri yang disesuaikan dengan karakteristik masing-masing platform, beberapa contoh kueri yang digunakan antara lain:

Google, site:.id intext:"slot"

Censys, web.endpoints.http.html_title: "slot" AND web.hostname: ".sch.id"

Shodan, hostname:.go.id title:"slot"

Fofa, body="judi online gacor" && country="ID"

Hasil pencarian dari masing-masing platform menghasilkan daftar domain yang kemudian dijadikan sebagai data pembandingan akurasi yang selanjutnya akan diproses oleh *web scraper*, Tabel 8 berikut merangkum hasil temuan.

Tabel 8. Hasil Sumber Data

Platform	Domain	Terindikasi	Server	Tanggal Temuan
Google	www.mediastartup.id	Apa itu Slot Gacor Online? Slot Gacor <i>Online</i> adalah istilah untuk permainan slot yang memiliki tingkat kemenangan tinggi	-	2019
Censys	rtp.maalmusthofa.sch.id	RTP LIVE > Situs Bocoran Game Slot Gacor Terbaru Hari Ini Gampang Maxwin	F5 Nginx	10 Nov 2025
Shodan	guruinspiratif.kemdikbud.go.id	SLOT88: 21 Daftar Situs Slot Gacor Terpercaya Gampang Menang Hari Ini	nginx/1.16.1	2025-11-10
Fofa	bpmi.unucirebon.ac.id	SLOT777 : Situs Toto Slot Gacor Bet 200 Malam Hari Ini Pasti Maxwin	Apache	2025-11-07

3.4. Hasil Akurasi

Sebelum dilakukan pengujian akurasi secara formal terhadap 50 domain, *web scraper* telah melewati beberapa tahap pengujian awal dengan total kurang lebih 40 domain dari berbagai variasi struktur situs. pada fase uji awal tersebut, sistem menunjukkan adanya beberapa kendala deteksi, yaitu munculnya kesalahan *false positive* sebanyak 3 kasus dan *false negative* sebanyak 2 kasus. kesalahan ini terutama terjadi pada domain dengan pola penulisan kata yang ambigu atau halaman yang memuat tautan eksternal tersembunyi sehingga sistem belum mampu membedakannya secara optimal. temuan pada tahap uji 40 domain ini kemudian digunakan sebagai dasar untuk melakukan penyempurnaan pada mekanisme pencarian kata kunci, pola frasa

kontekstual, serta logika *traversal* DFS agar lebih stabil. setelah perbaikan tersebut diterapkan, dilakukan kembali pengujian akurasi menggunakan 50 domain sampel.

Berikut Tabel 9 menunjukkan rekapitulasi hasil pengujian yang dilakukan terhadap seluruh domain uji yang dipilih.

Tabel 9. Rekapitulasi Hasil Pengujian Akurasi

Keterangan	Jumlah Kasus	Persentase
Analisis benar (<i>True Positive + True Negative</i>)	49	98%
Salah analisis (<i>False Positive + False Negative</i>)	1	2%
Total Pengujian	50	100%

Rumus yang digunakan akurasi sistem, berdasarkan hasil pengujian diatas dengan perhitungan berikut:

$$\text{Akurasi} = \frac{33 + 16}{33 + 16 + 1 + 0} \times 100\%$$

$$\text{Akurasi} = \frac{49}{50} \times 100\%$$

$$\text{Akurasi} = 98\%$$

Secara keseluruhan, hasil akurasi ini menunjukkan bahwa *web scraper* yang dikembangkan telah mampu melakukan analisis penyusupan *judi online* dengan tingkat ketepatan yang cukup tinggi, dan hasil ini akan digunakan sebagai dasar dalam proses evaluasi lanjutan berikutnya.

Evaluasi dilakukan untuk menilai kinerja sistem *web scraper* berdasarkan tiga indikator utama, yaitu ketepatan hasil deteksi, ketepatan domain, dan jumlah kesalahan deteksi. Hasil evaluasi disajikan pada Tabel 10 berikut.

Tabel 10. Evaluasi Hasil *Web Scraper*

Hasil Pengamatan	Tingkat Keberhasilan
Dari 50 domain yang diuji, 49 hasil sesuai dan 1 kurang sesuai.	98%
Sistem tidak berpindah domain selama proses <i>scraper</i> .	100%
Ditemukan 1 kurang tepatnya analisis pada FP.	2%

Sebagai bagian dari proses evaluasi lanjutan, dilakukan kembali pengujian dengan cakupan yang lebih besar, yaitu sebanyak 60 domain tambahan untuk memastikan konsistensi kinerja sistem pada kondisi yang lebih beragam. pada tahap evaluasi ini, sistem sempat menghasilkan 3 kasus *false positive*, terutama pada domain yang memiliki kemiripan pola kata dengan istilah perjudian tetapi sebenarnya tidak berkaitan. berdasarkan temuan tersebut, dilakukan penyempurnaan terhadap modul analisis konteks, terutama pada proses identifikasi frasa dan pemeriksaan tautan eksternal. setelah perbaikan diterapkan, pengujian ulang menunjukkan peningkatan kinerja yang signifikan, dengan jumlah *false positive* menurun menjadi hanya 1 kasus. penurunan kesalahan ini berdampak langsung pada kenaikan tingkat akurasi keseluruhan menjadi 98,33%. hasil evaluasi lanjutan ini memperkuat bukti bahwa penyempurnaan mekanisme analisis dan penelusuran berbasis DFS mampu meningkatkan ketepatan sistem dalam mendeteksi penyusupan *link judi online*.

$$\text{Akurasi} = \frac{42 + 17}{42 + 17 + 1 + 0} \times 100\%$$

$$\text{Akurasi} = \frac{59}{60} \times 100\%$$

$$\text{Akurasi} = 98,33\%$$

Meskipun perhitungan awal menunjukkan tingkat akurasi sebesar 98%, evaluasi lanjutan terhadap satu kasus anomali di mana sistem sempat mendeteksi keberadaan *judi online* yang sebenarnya tidak tepat bahwa kesalahan tersebut bukan disebabkan oleh kelemahan logika atau metode deteksi *scraper*. setelah dilakukan pemeriksaan manual, kasus tersebut dikonfirmasi sebagai FP *False Positive* yang terisolasi, sehingga akurasi operasional sistem tetap dapat dikategorikan mendekati 100%. hal ini menegaskan bahwa *scraper* yang dikembangkan memiliki performa deteksi yang kuat, meskipun tetap perlu pemahaman bahwa tidak ada sistem otomatis yang sepenuhnya bebas dari kesalahan.

4. KESIMPULAN

Berdasarkan hasil penelitian mengenai pengembangan sistem *web scraper* untuk menganalisis penyusupan situs *judi online* pada domain ac.id, go.id, my.id, .id, dan sch.id, dapat disimpulkan bahwa:

1. Sistem *web scraper* yang dikembangkan mampu melakukan proses *crawling* menggunakan algoritma *Depth First Search* pada lima domain target. sistem berhasil menelusuri struktur halaman, mengumpulkan tautan, serta mengekstraksi elemen

- seperti teks dan tag html yang berkaitan dengan penyusupan situs *judi online*. kinerja *crawling* berjalan sesuai alur DFS dan menunjukkan bahwa cara ini efektif digunakan dalam proses penelusuran mendalam pada struktur website yang kompleks.
- Hasil pengujian akurasi menunjukkan bahwa sistem *web scraper* memiliki ketepatan deteksi dalam menganalisis penyusupan situs *judi online*, pada tahap uji awal terhadap 40 domain, sistem masih menghasilkan beberapa kesalahan berupa *false positive* dan *false negative*. temuan tersebut kemudian digunakan sebagai dasar penyempurnaan pada mekanisme *web scraper*. setelah perbaikan diterapkan, pengujian terhadap 50 domain menunjukkan akurasi sebesar 98%, dengan 49 analisis sesuai dan hanya 1 kesalahan deteksi. evaluasi lanjutan pada 60 domain tambahan kembali memperlihatkan peningkatan konsistensi kinerja, di mana jumlah kesalahan *false positive* berhasil menjadi 1 kasus setelah penyempurnaan modul analisis, sehingga akurasi meningkat menjadi 98,33%. hasil ini menegaskan bahwa mekanisme deteksi berbasis analisis dan *crawling* DFS pada sistem yang dikembangkan mampu bekerja secara baik.

DAFTAR PUSTAKA

- [1] “Implementasi Business Intelligence Menggunakan Tableau Untuk Visualisasi Data Dampak Judi Online Di Indonesia,” *Jurnal Ilmiah Komputasi*, vol. 23, no. 2, Jun. 2024, doi: 10.32409/jikstik.23.2.3609.
- [2] M. Nurseno, U. Aditiawarman, H. Al Qodri Maarif, and T. Mantoro, “Detecting Hidden Illegal Online Gambling on .go.id Domains Using Web Scraping Algorithms,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 2, pp. 365–378, Mar. 2024, doi: 10.30812/matrik.v23i2.3824.
- [3] D. Hasiyanto Hendarto, P. Kejahatan Siber Terkait Distribusi Perjudian Online di Indonesia dalam Rangka Mewujudkan Keamanan dan Ketertiban Masyarakat, and R. Sri Handayani, “How to cite: Pencegahan Kejahatan Siber Terkait Distribusi Perjudian Online di Indonesia dalam Rangka Mewujudkan Keamanan dan Ketertiban Masyarakat,” vol. 5, no. 5, p. p-ISSN, 2024.
- [4] S. Saeed, L. Fotia, M. Javed, M. Chowdhury, and M. A. Tamal, “Dataset of suspicious phishing URL detection,” 2024, doi: 10.17632/6tm2d6sz7p.1.
- [5] Sekarwangi, S.N., et al. 2024. “Penanggulangan Fenomena Judi Online Melalui Pendekatan Media Komik Digital.” *Communicare: Journal of Communication Studies*, 6(1), 41–52.
- [6] N. Gul, N. A. Megha K Sathyan, and M. Dnyanesh Bharat, “WEB SCRAPER FOR DATA EXTRACTION AND THREAT INTELLIGENCE.”
- [7] I. And and D. Expert, “Pengembangan Aplikasi Scanning Defacement Judi Online untuk Website Profile pada Server INFORMASI ARTIKEL A B S T R A K,” 2024. [Online]. Available: <http://e-journal.unper.ac.id>
- [8] B. Wibowo, A. Fathl Jannah, and L. Hafiz, “Optimalisasi Bot Telegram untuk Deteksi Situs Perjudian Online di Dunia Pendidikan dan Sektor Pemerintah,” *Jurnal Pengabdian Masyarakat Sultan Indonesia*, vol. 2, no. 1, pp. 17–24, Dec. 2024, doi: 10.58291/abdisultan.v2i1.316.
- [9] Raharja, Y. 2024. “Implementasi Metode OSINT untuk Mengidentifikasi Serangan Judi Online pada Website.” *Jurnal Informatika Polinema*, 10(3), 359–364.
- [10] Rachmatika, A.N. 2024. Penerapan Black Box Testing dan Pendekatan Behavior Driven Development pada Sistem Ujian Computer Based Test. Tugas Akhir Tidak Diterbitkan. Malang: Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
- [11] M. Fadli Mutaqin, D. Ferdiansyah, J. Setiabudhi no, and J. Barat, “Identifikasi Kerentanan Terhadap Serangan Slot Backdoor Pada Website di Indonesia Dengan Menggunakan Metode OSINT,” vol. 1, no. 2, pp. 2986–5360, doi: 10.29322/IJSRP.X.X.2018.pXXXX.
- [12] H. Mukhtar, J. Al Amien, and M. A. Rucyat, ‘Filtering Spam Email menggunakan Algoritma Naïve Bayes’, *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 1, pp. 9–19, May 2022, doi: 10.37859/coscitech.v3i1.3652.
- [13] A. Oktian Permana and Sudin Saepudin, ‘Perbandingan algoritma k-nearest neighbor dan naïve bayes pada aplikasi shopee’, *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 25–32, Apr. 2023, doi: 10.37859/coscitech.v4i1.4474.