Jurnal Publikasi Teknik Informatika
Volume 5 Nomor 3, September 2026
e-ISSN: 2808-8972 p-ISSN: 2808-9367, Hal. 92-110
DOI: https://doi.org/10.55606/jupti.v5i3.7994
Tersedia: https://journalcenter.org/index.php/jupti

Deteksi Domain DGA menggunakan Random Forest Berdasarkan Karakteristik String dengan Evaluasi Unseen Family

Muhammad Davin Putra Pratama Arisandi[1], Arief Akhmad Fauzi[2*]
[1-2]Program Studi Informatika, Institut Teknologi Nasional Bandung, Indonesia
Email: davinarisandi93@gmail.com[1], ariefaf.univ@gmail.com[2]
**Penulis Korespondensi: ariefaf.univ@gmail.com*

Abstract*. Domain Generation Algorithm (DGA) is a mechanism used by malware to algorithmically generate domain names as candidate destinations for communication with command and control servers. Domain string characteristics can be used to distinguish algorithmically generated domains from legitimate domains. This study aims to identify DGA domains based on string characteristics using Random Forest and evaluate its performance against DGA families not observed during training. The dataset consists of 20,000 balanced domains, comprising 10,000 DGA domains from 52 families and 10,000 legitimate domains. A total of 21 features were extracted, including domain length, name length, character counts and ratios, unique character counts, domain structure, and Shannon entropy. Evaluation was conducted using Random Stratified Split and Unseen Family Split. Random Stratified Split achieved 91.83% accuracy, 92.14% precision, 91.45% recall, and 91.79% F1-score. Unseen Family Split achieved 85.55% accuracy, 91.47% precision, 78.41% recall, and 84.44% F1-score. The results indicate a change in classification performance when the model encounters unseen DGA families. Feature importance identified name length and the ratio of name length to domain length as the two most important features. The findings provide an evaluation basis for DGA classification under unseen-family conditions.*

Keywords*: Characteristic String; DGA Detection; Domain Generation Algorithm; Random Forest; Unseen Family.*

Abstrak. *Domain Generation Algorithm* merupakan mekanisme yang digunakan malware untuk menghasilkan nama domain secara algoritmik sebagai kandidat komunikasi dengan server *command and control*. Karakteristik string domain dapat digunakan sebagai informasi untuk membedakan domain yang dihasilkan secara algoritmik dan domain *legitimate*. Penelitian ini bertujuan mengidentifikasi domain *Domain Generation Algorithm* berdasarkan karakteristik string menggunakan *Random Forest* serta mengevaluasi kemampuan model terhadap *DGA family* yang tidak diamati selama pelatihan. Dataset penelitian terdiri atas 20.000 domain yang seimbang, yaitu 10.000 domain DGA dari 52 *family* dan 10.000 domain *legitimate*. Sebanyak 21 fitur diekstraksi dari karakteristik string domain, meliputi panjang domain, panjang nama domain, jumlah dan rasio karakter, jumlah karakter unik, struktur domain, serta *Shannon entropy*. Evaluasi dilakukan menggunakan *Random Stratified Split* dan *Unseen Family Split*. Pada *Random Stratified Split*, model menghasilkan *accuracy* 91,83%, *precision* 92,14%, *recall* 91,45%, dan *F1-score* 91,79%. Pada *Unseen Family Split*, diperoleh *accuracy* 85,55%, *precision* 91,47%, *recall* 78,41%, dan *F1-score* 84,44%. Pengujian *Unseen Family Split* menggunakan delapan *DGA family* yang tidak memiliki *overlap* dengan *family* pada data pelatihan. Hasil pengujian menunjukkan perubahan kinerja ketika model menghadapi *DGA family* yang tidak digunakan selama pelatihan, sedangkan *feature importance* menunjukkan *name length* dan rasio nama domain terhadap panjang domain sebagai dua fitur dengan kontribusi terbesar. Temuan tersebut memberikan dasar evaluasi klasifikasi DGA pada kondisi ketika *DGA family* yang dihadapi tidak terdapat dalam data pelatihan.

Kata Kunci: Deteksi DGA; *Domain Generation Algorithm*; Karakteristik String; *Random Forest*; *Unseen Family.*

1. LATAR BELAKANG

Internet dan layanan digital yang semakin luas menjadikan keamanan komunikasi jaringan penting untuk melindungi pengguna dan infrastruktur dari aktivitas berbahaya. Salah satu sumber informasi untuk mengidentifikasi komunikasi terkait *malware* dan *command and control* (C2) adalah aktivitas *Domain Name System* (DNS). (Suryotrisongko, Musashi, Tsuneda, & Sugitani, 2022) menganalisis trafik DNS menggunakan karakteristik statistik untuk mendeteksi botnet berbasis *Domain Generation Algorithm* (DGA) dengan data yang mencakup

Naskah Masuk: 22 Juni 2026; Revisi: 25 Juli 2026; Diterima: 29 Agustus 2026; Terbit: 04 September 2026

55 *DGA family*. DGA menghasilkan nama domain secara algoritmik sebagai kandidat komunikasi dengan server C2, sehingga perubahan domain secara dinamis dapat membatasi mekanisme pertahanan berbasis daftar domain berbahaya (*blacklist*). (Wang, Guo, & Montgomery, 2022) menunjukkan bahwa malware berbasis DGA dapat menghasilkan sejumlah besar *algorithmically generated domains* (AGD), sedangkan (Suryotrisongko et al., 2022) menunjukkan bahwa karakteristik DNS dapat digunakan bersama *machine learning* untuk mendeteksi trafik DGA.

Karakteristik string domain dapat digunakan untuk membedakan domain DGA dan *legitimate*, antara lain melalui panjang, komposisi, struktur, dan keragaman karakter. (Wang et al., 2022) menggunakan fitur karakteristik domain dan *feature engineering* untuk mendeteksi AGD, sedangkan (Sun & Liu, 2023) mengombinasikan fitur domain, *Whois*, dan *N-gram* dengan *deep learning* untuk klasifikasi DGA. Penelitian lain mengembangkan pendekatan dengan karakteristik dan arsitektur yang beragam. (Liew & Law, 2023) membandingkan *machine learning* berbasis ekstraksi fitur dengan *deep learning* pada berbagai tipe DGA, sedangkan (Chen, Lang, Chen, & Xie, 2023) menggunakan *feature fusion* melalui segmentasi kata bermakna dan *N-gram*. (Suryotrisongko et al., 2022) mengintegrasikan karakteristik statistik trafik DNS dengan *explainable artificial intelligence* (XAI) dan *open-source intelligence* (OSINT). (Lee, Do Yoo, Jeong, & Kim, 2024) memperhatikan *false positive* pada domain berbahasa Mandarin, (Tang, Guan, Zhao, Wang, & Chen, 2024) menggunakan *Transformer* dan *Rapid Selective Kernel Network*, (Selvaraj & Panjanathan, 2024) mengkaji *word-level DGA* dan *adversarial DGA*, sedangkan (Biros & Kantor, 2025) mengevaluasi beberapa algoritma *machine learning*, termasuk *Random Forest*, pada *character-based DGA*.

Meskipun berbagai pendekatan telah digunakan, kemampuan generalisasi terhadap *DGA family* yang tidak diamati selama pelatihan masih perlu dievaluasi. (Liew & Law, 2023) menunjukkan adanya perbedaan hasil klasifikasi pada tipe DGA yang berbeda, sementara (Jeremiah et al., 2025) mengevaluasi model pada dataset yang mencakup lebih dari 16 juta domain dan 53 *DGA family* serta beberapa dataset eksternal untuk menguji generalisasi. Kondisi tersebut menunjukkan pentingnya mengevaluasi model ketika data pengujian berasal dari *DGA family* yang tidak digunakan selama pelatihan. Berdasarkan kondisi tersebut, penelitian ini mengidentifikasi domain DGA berdasarkan karakteristik string menggunakan *Random Forest* dan mengevaluasi perubahan kinerja model melalui *Random Stratified Split* dan *Unseen Family Split*. Sebanyak 21 fitur karakteristik string digunakan, meliputi panjang domain, panjang nama domain, komposisi dan rasio karakter, jumlah karakter unik, struktur

domain, serta *Shannon entropy*. Analisis *feature importance* dilakukan untuk mengidentifikasi kepentingan relatif fitur dalam klasifikasi.

2. KAJIAN TEORITIS

Domain Generation Algorithm

Domain Generation Algorithm (DGA) merupakan mekanisme yang digunakan malware untuk menghasilkan nama domain secara algoritmik sebagai kandidat tujuan komunikasi dengan server *command and control* (C2), dengan karakteristik string yang dapat berbeda dari domain *legitimate* sehingga dapat digunakan sebagai informasi dalam proses klasifikasi. (Wang et al., 2022) menggunakan karakteristik domain dan *feature engineering* dalam pendekatan *machine learning* untuk mendeteksi *algorithmically generated domains* (AGD), sedangkan (Sun & Liu, 2023) menggunakan karakteristik domain bersama informasi *Whois* dan *N-gram* untuk klasifikasi DGA. Hasil penelitian tersebut menjadi landasan penggunaan informasi yang diperoleh langsung dari string domain dalam penelitian ini.

Karakteristik String Domain

Karakteristik string domain digunakan sebagai representasi masukan dalam proses klasifikasi, yang mencakup panjang domain, panjang nama domain, jumlah dan rasio karakter, jumlah karakter unik, struktur domain, serta *Shannon entropy*. (Wang et al., 2022) menggunakan fitur berbasis karakteristik domain untuk mendeteksi AGD, sedangkan (Chen et al., 2023) mengombinasikan segmentasi kata bermakna dan *N-gram*, dan (Liew & Law, 2023) mengevaluasi representasi berbasis *subword tokenization* untuk klasifikasi DGA. (Yang, Lu, Yan, Zhang, & Yu, 2022) menggunakan kombinasi *N-gram* dan *Transformer* untuk menangkap pola kombinasi serta posisi karakter pada nama domain. Perbedaan bentuk representasi tersebut menunjukkan bahwa string domain dapat direpresentasikan melalui fitur eksplisit maupun representasi berbasis urutan karakter. Pada penelitian ini, karakteristik string direpresentasikan dalam 21 fitur sebagai masukan *Random Forest*, termasuk *Shannon entropy* untuk merepresentasikan tingkat keragaman karakter pada string domain.

Random Forest

Random Forest digunakan sebagai algoritma klasifikasi untuk membedakan domain DGA dan *legitimate* melalui sejumlah *decision tree* yang menghasilkan prediksi berdasarkan fitur masukan. (Biros & Kantor, 2025) menggunakan *Random Forest* sebagai salah satu algoritma *machine learning* untuk *character-based DGA*, sedangkan (Divya, Amritha, & Viswanathan, 2022) juga menerapkan Random Forest bersama beberapa algoritma klasifikasi untuk mendeteksi domain DGA berdasarkan karakteristik domain. Pada penelitian ini, model

dikonfigurasi menggunakan 300 *decision tree* melalui n_estimators=300, max_features=sqrt, random_state=42, dan n_jobs=-1, serta diterapkan secara konsisten pada kedua skenario eksperimen agar perubahan hasil evaluasi merefleksikan perbedaan pembagian data, bukan perubahan konfigurasi model.

Random Stratified Split dan Unseen Family Split

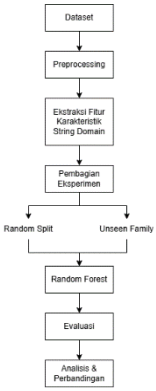
Evaluasi dilakukan melalui dua skenario pembagian data, yaitu *Random Stratified Split* yang mempertahankan proporsi kelas DGA dan *legitimate* pada data pelatihan dan pengujian, serta *Unseen Family Split* yang menempatkan sejumlah *DGA family* secara eksklusif pada data pengujian untuk mengevaluasi kinerja model terhadap pola domain yang tidak diamati selama pelatihan. (Peng, He, Ni, & Yang, 2026) juga mengevaluasi *novel DGA families* dengan menahan sejumlah *family* dari proses pelatihan dan menggunakannya sebagai data pengujian, sehingga pendekatan tersebut menjadi landasan penggunaan pemisahan *family* dalam penelitian ini.

Metrik Evaluasi

Kinerja klasifikasi dievaluasi menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan ROC-AUC, dengan *confusion matrix* untuk melihat distribusi prediksi benar dan salah pada kelas DGA dan *legitimate*. Selain itu, *feature importance* digunakan untuk mengidentifikasi kepentingan relatif setiap fitur dalam model *Random Forest* dengan mengurutkan nilai kepentingan dan memilih fitur dengan nilai tertinggi untuk divisualisasikan. Penggunaan beberapa metrik dan analisis *feature importance* memungkinkan evaluasi mencakup kinerja klasifikasi serta perbedaan kepentingan relatif fitur yang digunakan.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis eksperimen untuk mengidentifikasi domain *Domain Generation Algorithm* (DGA) berdasarkan karakteristik string menggunakan *Random Forest*. Pendekatan kuantitatif digunakan karena hasil pengujian dinyatakan dalam ukuran numerik berupa metrik klasifikasi dan nilai *feature importance*. Rancangan eksperimen meliputi penyusunan dataset, pra-pemrosesan data, ekstraksi fitur, pembentukan skenario eksperimen, pemodelan *Random Forest*, serta evaluasi dan perbandingan hasil klasifikasi. Tahapan penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian.

Gambar 1 menunjukkan tahapan penelitian yang dimulai dari penyusunan dataset, dilanjutkan dengan pra-pemrosesan dan ekstraksi 21 fitur karakteristik string domain. Data yang telah direpresentasikan dalam bentuk fitur kemudian dibagi menggunakan dua skenario eksperimen, yaitu *Random Stratified Split* dan *Unseen Family Split*. Kedua skenario digunakan sebagai masukan dalam pemodelan *Random Forest*, kemudian hasil prediksi dievaluasi dan dibandingkan berdasarkan metrik klasifikasi serta analisis *feature importance*. Pada *Unseen Family Split*, analisis tambahan dilakukan berdasarkan *recall* pada masing-masing *DGA family*.

Tahapan Penelitian

Tahapan penelitian diawali dengan penyusunan dataset yang terdiri atas domain DGA dan domain *legitimate*. Data kemudian melalui proses pra-pemrosesan yang mencakup normalisasi, validasi format, dan penghapusan data duplikat. Domain yang telah memenuhi kriteria selanjutnya digunakan untuk membentuk dataset eksperimen dengan jumlah kelas yang seimbang.

Tahap berikutnya adalah ekstraksi fitur karakteristik string domain. Setiap domain direpresentasikan ke dalam 21 fitur numerik yang digunakan sebagai variabel masukan pada model *Random Forest*. Setelah fitur terbentuk, eksperimen dilakukan menggunakan dua mekanisme pembagian data, yaitu *Random Stratified Split* dan *Unseen Family Split*.

Pada *Random Stratified Split*, pembagian data dilakukan secara acak dengan mempertahankan proporsi kelas. Sementara itu, *Unseen Family Split* menggunakan *DGA family* sebagai dasar pembentukan kelompok sehingga sejumlah *family* tidak digunakan dalam proses pelatihan. Hasil prediksi dari kedua skenario kemudian dievaluasi menggunakan metrik klasifikasi. Perbandingan hasil digunakan untuk mengamati perubahan kinerja model pada kedua kondisi pengujian.

Dataset Penelitian

Dataset penelitian terdiri atas dua kelas, yaitu domain DGA dan domain *legitimate*. Data DGA merupakan kumpulan domain yang dilengkapi informasi *DGA family*, sedangkan data *legitimate* berasal dari kumpulan domain populer. Informasi *DGA family* pada data DGA dipertahankan selama proses pengolahan karena digunakan sebagai dasar pembentukan skenario *Unseen Family Split*.

Setelah proses pembersihan data, diperoleh 1.379.809 domain DGA dari 52 *DGA family* dan 694.778 domain *legitimate*. Dari kumpulan tersebut, masing-masing kelas diambil sebanyak 10.000 domain menggunakan *random state* 42 sehingga diperoleh dataset penelitian sebanyak 20.000 domain dengan distribusi kelas yang seimbang. Komposisi dataset penelitian ditunjukkan pada Tabel 1.

Tabel 1. Komposisi dataset penelitian.

Kelas	Jumlah	Label
DGA	10.000	1
Legitimate	10.000	0
Total	**20.000**	—

Berdasarkan Tabel 1, kedua kelas memiliki jumlah data yang sama sehingga dataset eksperimen memiliki distribusi kelas 1:1. Informasi *DGA family* tetap dipertahankan pada data DGA untuk digunakan sebagai kelompok pada skenario *Unseen Family Split*.

Pra-pemrosesan Data

Pra-pemrosesan dilakukan sebelum ekstraksi fitur untuk menyeragamkan representasi domain dan memastikan data memenuhi kriteria pengolahan. Normalisasi dilakukan dengan mengubah karakter domain menjadi huruf kecil, menghilangkan awalan protokol HTTP/HTTPS dan *www*, mengambil bagian domain yang relevan, serta menghilangkan titik pada akhir domain.

Validasi dilakukan dengan memeriksa keberadaan nama domain, panjang maksimum domain sebesar 253 karakter, pemisah antarlabel, karakter spasi, dan pola karakter yang diperbolehkan. Setiap *label* domain juga diperiksa agar tidak kosong, tidak diawali atau diakhiri tanda hubung, serta memiliki panjang maksimum 63 karakter.

Setelah proses validasi, data duplikat dihapus dari masing-masing kelompok data. Domain *legitimate* yang memiliki kesamaan dengan domain pada kelompok DGA juga dikeluarkan untuk mencegah domain yang sama digunakan pada kedua kelas. Data yang telah melalui proses tersebut kemudian digunakan untuk membentuk dataset eksperimen sebanyak 10.000 domain DGA dan 10.000 domain *legitimate*.

Ekstraksi Fitur Karakteristik String

Setiap domain direpresentasikan ke dalam bentuk numerik berdasarkan karakteristik string. Ekstraksi fitur dilakukan setelah proses normalisasi dengan memisahkan nama domain dan *top-level domain* (TLD), kemudian menghitung karakteristik string dari domain yang telah diproses. Proses tersebut menghasilkan 21 fitur yang digunakan sebagai variabel masukan model. Fitur karakteristik string domain ditunjukkan pada Tabel 2.

Tabel 2. Fitur karakteristik string domain.

No.	Fitur	Kelompok
1	Domain length	Panjang
2	Name length	Panjang
3	TLD length	Struktur
4	Label count	Struktur
5	Subdomain count	Struktur
6	Letter count	Komposisi
7	Digit count	Komposisi
8	Hyphen count	Komposisi
9	Dot count	Struktur
10	Unique character count	Keragaman
11	Vowel count	Komposisi
12	Consonant count	Komposisi
13	Letter ratio	Rasio
14	Digit ratio	Rasio
15	Vowel ratio	Rasio
16	Consonant ratio	Rasio
17	Unique character ratio	Keragaman
18	Name-domain ratio	Rasio
19	Shannon entropy	Entropi
20	Has digit	Indikator
21	Has hyphen	Indikator

Tabel 2 menunjukkan 21 fitur yang digunakan untuk merepresentasikan karakteristik string domain. Fitur panjang terdiri atas *domain length* dan *name length*. Fitur komposisi mencakup jumlah huruf, angka, tanda hubung, vokal, dan konsonan. Fitur rasio dihitung berdasarkan jumlah karakter terhadap panjang string yang relevan, sedangkan fitur keragaman merepresentasikan jumlah dan proporsi karakter unik.

Fitur struktur mencakup panjang TLD, jumlah *label*, jumlah *subdomain*, dan jumlah titik. *Name-domain ratio* merepresentasikan perbandingan panjang nama domain terhadap panjang keseluruhan domain. *Shannon entropy* dihitung berdasarkan distribusi karakter pada string domain tanpa tanda titik. Dua fitur indikator, yaitu *has digit* dan *has hyphen*, masing-masing menunjukkan keberadaan angka dan tanda hubung pada domain.

Skenario Eksperimen

Penelitian menggunakan dua skenario eksperimen untuk mengevaluasi model pada kondisi pembagian data yang berbeda. Kedua skenario menggunakan dataset dan 21 fitur yang sama, sedangkan mekanisme pembagian data dibedakan berdasarkan tujuan evaluasi.

Random Stratified Split

Pada skenario *Random Stratified Split*, dataset dibagi menjadi data pelatihan dan pengujian dengan proporsi 80:20. Pembagian dilakukan secara acak dengan *test size* sebesar 20% dan *random state* 42, sedangkan *stratification* diterapkan berdasarkan label kelas.

Pembagian tersebut menghasilkan 16.000 data pelatihan dan 4.000 data pengujian. Data pelatihan terdiri atas 8.000 domain DGA dan 8.000 domain *legitimate*, sedangkan data pengujian terdiri atas 2.000 domain DGA dan 2.000 domain *legitimate*. Dengan demikian, proporsi kelas tetap seimbang pada data pelatihan dan pengujian.

Skenario ini digunakan untuk memperoleh hasil klasifikasi ketika domain pada data pelatihan dan pengujian dibentuk melalui pembagian secara acak dengan distribusi kelas yang dipertahankan.

Unseen Family Split

Pada skenario *Unseen Family Split*, *DGA family* digunakan sebagai unit kelompok dalam proses pembagian data. Pembagian dilakukan dengan *test size* sebesar 20% dan *random state* 42. Sejumlah kandidat pembagian dibentuk, kemudian dipilih pembagian yang menghasilkan jumlah data pengujian paling mendekati target 20%.

Data DGA hasil pembagian kemudian diseimbangkan dengan data *legitimate* menggunakan jumlah data yang sama pada masing-masing kelompok. Hasil pembagian menghasilkan 15.952 data pelatihan dan 4.048 data pengujian. Data pelatihan terdiri atas 7.976 domain DGA dan 7.976 domain *legitimate*, sedangkan data pengujian terdiri atas 2.024 domain DGA dan 2.024 domain *legitimate*.

Pembagian tersebut menghasilkan 31 *DGA family* pada data pelatihan dan 8 *DGA family* pada data pengujian, tanpa *overlap* antara kedua kelompok. Kondisi tersebut diverifikasi dengan memeriksa irisan *DGA family* pada data pelatihan dan pengujian. Perbandingan konfigurasi kedua skenario ditunjukkan pada Tabel 3.

Tabel 3. Konfigurasi skenario eksperimen.

Parameter	Random Stratified Split	Unseen Family Split
Data pelatihan	16.000	15.952
Data pengujian	4.000	4.048
DGA pelatihan	8.000	7.976
DGA pengujian	2.000	2.024
Legitimate pelatihan	8.000	7.976
Legitimate pengujian	2.000	2.024
DGA family pelatihan	52*	31
DGA family pengujian	—	8
Family overlap	—	0

Keterangan: pada Random Stratified Split, pembagian dilakukan berdasarkan domain secara acak sehingga jumlah family tidak digunakan sebagai dasar pembentukan kelompok.

Tabel 3 menunjukkan bahwa perbedaan utama kedua skenario terletak pada unit pembagian data. *Random Stratified Split* mempertahankan distribusi kelas melalui pembagian domain secara acak, sedangkan *Unseen Family Split* memisahkan *DGA family* tertentu dari proses pelatihan sehingga delapan *family* hanya terdapat pada data pengujian.

Pemodelan dan Evaluasi

Pemodelan dilakukan menggunakan *Random Forest* dengan 21 fitur karakteristik string sebagai variabel masukan. Model dilatih dan diuji secara terpisah pada *Random Stratified Split* dan *Unseen Family Split*. Hasil prediksi kemudian dievaluasi menggunakan metrik klasifikasi, *confusion matrix*, dan ROC-AUC. Analisis *feature importance* dilakukan untuk memperoleh nilai kepentingan relatif setiap fitur. Pada *Unseen Family Split*, evaluasi tambahan dilakukan berdasarkan *recall* pada setiap *DGA family*.

Pemodelan Random Forest

Random Forest digunakan sebagai model klasifikasi dengan 21 fitur karakteristik string sebagai variabel masukan. Model dilatih secara terpisah menggunakan data pelatihan dari masing-masing skenario eksperimen.

Model dikonfigurasi menggunakan 300 *decision tree*. Parameter *max features* menggunakan nilai *square root* dari jumlah fitur, sedangkan *random state* ditetapkan sebesar 42 untuk menjaga keterulangan hasil eksperimen. Proses komputasi dijalankan menggunakan seluruh *logical processor* yang tersedia.

Model menghasilkan prediksi kelas untuk setiap domain pada data pengujian. Selain prediksi kelas, model menghasilkan probabilitas kelas DGA yang digunakan dalam perhitungan ROC-AUC dan pembentukan kurva ROC.

Metrik Evaluasi

Evaluasi dilakukan menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan *Receiver Operating Characteristic-Area Under the Curve* (ROC-AUC). *Confusion matrix* digunakan untuk menunjukkan distribusi prediksi benar dan salah pada kelas DGA dan *legitimate*. Seluruh metrik dihitung berdasarkan label aktual dan hasil prediksi model.

Accuracy digunakan untuk menghitung proporsi seluruh data yang diklasifikasikan dengan benar. Perhitungan *accuracy* ditunjukkan pada Persamaan (i).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{(i)}$$

dengan TP (*true positive*) merupakan jumlah domain DGA yang diprediksi sebagai DGA, TN (*true negative*) merupakan jumlah domain *legitimate* yang diprediksi sebagai *legitimate*, FP (*false positive*) merupakan jumlah domain *legitimate* yang diprediksi sebagai

DGA, dan FN (*false negative*) merupakan jumlah domain DGA yang diprediksi sebagai *legitimate*.

Selanjutnya, *precision* digunakan untuk menghitung proporsi prediksi DGA yang benar terhadap seluruh data yang diprediksi sebagai DGA. Perhitungan *precision* ditunjukkan pada Persamaan (ii).

$$Precision = \frac{TP}{TP+FP} \quad \text{(ii)}$$

dengan TP merupakan jumlah domain DGA yang diprediksi sebagai DGA dan FP merupakan jumlah domain *legitimate* yang diprediksi sebagai DGA.

Recall digunakan untuk menghitung proporsi domain DGA yang berhasil terdeteksi dari seluruh domain DGA pada data pengujian. Perhitungan *recall* ditunjukkan pada Persamaan (iii).

$$Recall = \frac{TP}{TP+FN} \quad \text{(iii)}$$

dengan TP merupakan jumlah domain DGA yang diprediksi sebagai DGA dan FN merupakan jumlah domain DGA yang diprediksi sebagai *legitimate*.

F1-score digunakan untuk menggabungkan *precision* dan *recall* melalui rata-rata harmonis. Perhitungannya ditunjukkan pada Persamaan (iv).

$$F1 - score = 2\,x\,\frac{Precision\ x\ Recall}{Precision+Recall} \quad \text{(iv)}$$

Dengan *Precision* merupakan nilai *precision* dan *Recall* merupakan nilai *recall* yang diperoleh dari hasil klasifikasi.

ROC-AUC digunakan untuk mengevaluasi kemampuan model dalam membedakan kelas berdasarkan probabilitas prediksi pada berbagai nilai ambang. Nilai AUC diperoleh menggunakan roc_auc_score, sedangkan kurva ROC dibentuk menggunakan roc_curve.

Pada skenario *Unseen Family Split*, evaluasi tambahan dilakukan dengan menghitung *recall* untuk setiap *DGA family*. Perhitungan dilakukan berdasarkan jumlah domain pada setiap *family* yang diprediksi sebagai DGA dibandingkan dengan seluruh domain DGA pada *family* tersebut. Selain metrik klasifikasi, *feature importance* dari *Random Forest* digunakan untuk memperoleh nilai kepentingan relatif setiap fitur. Sepuluh fitur dengan nilai *importance* tertinggi kemudian digunakan dalam visualisasi untuk masing-masing skenario eksperimen.

4. HASIL DAN PEMBAHASAN

Hasil penelitian disajikan berdasarkan tahapan eksperimen yang meliputi pembentukan dataset, pengujian menggunakan *Random Stratified Split*, pengujian menggunakan *Unseen Family Split*, perbandingan kinerja kedua skenario, analisis *feature importance*, dan evaluasi *recall* pada *DGA family* yang tidak digunakan selama pelatihan. Penyajian tersebut digunakan untuk mengamati perubahan kinerja *Random Forest* ketika kondisi pembagian data berbeda.

Pembentukan Dataset dan Skenario Eksperimen

Dataset penelitian terdiri atas 20.000 domain dengan distribusi kelas seimbang, yaitu 10.000 domain DGA dan 10.000 domain *legitimate*. Data DGA mencakup 52 *DGA family*, sedangkan domain *legitimate* digunakan sebagai kelas pembanding. Setiap domain direpresentasikan menggunakan 21 fitur karakteristik string yang diperoleh melalui proses ekstraksi fitur.

Eksperimen dilakukan menggunakan dua skenario, yaitu *Random Stratified Split* dan *Unseen Family Split*. Pembagian data pada kedua skenario ditunjukkan pada Tabel 4.

Tabel 4. Pembagian dataset pada dua skenario eksperimen.

Skenario	Data Pelatihan	Data Pengujian	DGA Pengujian	DGA Family Pengujian
Random Stratified Split	16.000	4.000	2.000	Tidak dikontrol
Unseen Family Split	15.952	4.048	2.024	8

Berdasarkan Tabel 4, *Random Stratified Split* menghasilkan 16.000 data pelatihan dan 4.000 data pengujian dengan distribusi kelas yang dipertahankan. Pada *Unseen Family Split*, diperoleh 15.952 data pelatihan dan 4.048 data pengujian. Delapan *DGA family* ditempatkan secara eksklusif pada data pengujian sehingga tidak terdapat *family overlap* antara data pelatihan dan pengujian.

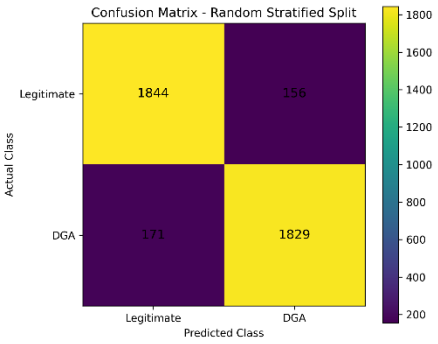
Pemisahan berdasarkan *DGA family* memberikan kondisi evaluasi yang berbeda dari pembagian data secara acak. (Peng et al., 2026) juga menggunakan pemisahan beberapa *DGA family* dari data pelatihan untuk mengevaluasi kemampuan model terhadap *novel DGA families*.

Selain perbedaan mekanisme pembagian data, variasi distribusi *DGA family* juga menjadi perhatian dalam penelitian deteksi DGA. (Fan, Ma, Liu, Yuan, & Ke, 2024) membahas permasalahan *long-tailed DGA detection*, yaitu perbedaan jumlah sampel antar-*family* yang dapat menyebabkan kinerja klasifikasi berbeda pada kategori dengan jumlah data yang kecil.

Kondisi tersebut menjadi pertimbangan dalam interpretasi hasil *recall* pada setiap *DGA family* dalam penelitian ini.

Hasil Random Stratified Split

Pengujian pertama dilakukan menggunakan *Random Stratified Split* dengan 4.000 data pengujian yang terdiri atas 2.000 domain DGA dan 2.000 domain *legitimate*. Hasil prediksi model terhadap kedua kelas digunakan untuk membentuk *confusion matrix*. Hasil *confusion matrix* pada skenario ini ditampilkan pada Gambar 2.



Gambar 2. Confusion matrix pada Random Stratified Split.

Berdasarkan Gambar 2, sebanyak 1.844 domain *legitimate* diklasifikasikan dengan benar sebagai *legitimate*, sedangkan 156 domain diklasifikasikan sebagai DGA. Pada kelas DGA, sebanyak 1.829 domain diklasifikasikan dengan benar sebagai DGA dan 171 domain diklasifikasikan sebagai *legitimate*. Dengan demikian, terdapat 3.673 prediksi benar dari 4.000 data pengujian. Nilai evaluasi pada kedua skenario ditampilkan pada Tabel 5.

Tabel 5. Hasil evaluasi Random Forest.

Skenario	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC
Random Stratified Split	91,83	92,14	91,45	91,79	0,9653
Unseen Family Split	85,55	91,47	78,41	84,44	0,9066

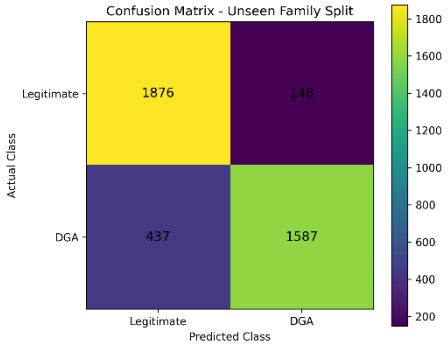
Pada *Random Stratified Split*, model menghasilkan *accuracy* 91,83%, *precision* 92,14%, *recall* 91,45%, *F1-score* 91,79%, dan ROC-AUC 0,9653. Hasil tersebut diperoleh ketika pembagian data dilakukan secara acak dengan distribusi kelas yang dipertahankan.

Penggunaan karakteristik string sebagai masukan klasifikasi memiliki kesesuaian dengan penelitian terdahulu. (Vranken & Alizadeh, 2022) menggunakan representasi TF-IDF berbasis *n-gram* untuk mendeteksi domain DGA, sedangkan penelitian ini menggunakan 21 fitur karakteristik string yang diekstraksi secara eksplisit dari domain. Pendekatan berbasis fitur juga digunakan oleh (Nie et al., 2023), yang mengombinasikan fitur buatan dengan model *machine learning*, termasuk *Random Forest*, dalam deteksi domain DGA. Pendekatan berbasis representasi domain juga digunakan oleh (Xie, He, & He, 2024), 2024), yang menggabungkan

CNN, Transformer, dan BiLSTM melalui *multi-stage feature fusion* untuk mendeteksi domain DGA. Model tersebut menghasilkan *accuracy* rata-rata 93,26% dan *F1-score* 93,32% pada 33 *DGA family*. Perbandingan tersebut tidak digunakan untuk menyatakan model yang lebih unggul karena dataset, jumlah *family*, representasi fitur, dan arsitektur model berbeda.

Hasil Unseen Family Split

Pengujian berikutnya dilakukan menggunakan *Unseen Family Split* untuk mengevaluasi model ketika menghadapi *DGA family* yang tidak digunakan selama pelatihan. Data pengujian terdiri atas 4.048 domain, yaitu 2.024 domain DGA dan 2.024 domain *legitimate*. Delapan *DGA family* ditempatkan secara eksklusif pada kelompok pengujian. Distribusi hasil prediksi ditampilkan pada Gambar 3.



Gambar 3. Confusion matrix pada Unseen Family Split.

Berdasarkan Gambar 3, sebanyak 1.876 domain *legitimate* diklasifikasikan dengan benar sebagai *legitimate*, sedangkan 148 domain diklasifikasikan sebagai DGA. Pada kelas DGA, 1.587 domain berhasil diklasifikasikan sebagai DGA dan 437 domain diklasifikasikan sebagai *legitimate*.

Berdasarkan Tabel 5, *Unseen Family Split* menghasilkan *accuracy* 85,55%, *precision* 91,47%, *recall* 78,41%, *F1-score* 84,44%, dan ROC-AUC 0,9066. Dibandingkan dengan *Random Stratified Split*, perubahan terbesar terdapat pada *recall*.

Recall yang lebih rendah menunjukkan bahwa sebagian domain DGA dari *family* yang tidak digunakan selama pelatihan lebih banyak diklasifikasikan sebagai *legitimate*. Hal tersebut terlihat dari peningkatan *false negative* menjadi 437 domain pada *Unseen Family Split*, dibandingkan 171 domain pada *Random Stratified Split*.

Hasil tersebut berkaitan dengan evaluasi generalisasi terhadap *DGA family* yang tidak diamati selama pelatihan. (Peng et al., 2026) juga menggunakan pengujian terhadap *novel DGA families* untuk mengamati kemampuan model di luar karakteristik yang tersedia selama proses pelatihan.

Pendekatan deteksi DGA dengan arsitektur berbeda juga menunjukkan bahwa karakteristik dan representasi domain memengaruhi hasil klasifikasi. (Jiang, Wu, Huang, & Deng, 2025), misalnya, menggunakan *gated convolution* dan LSTM untuk menangkap karakteristik lokal dan hubungan kontekstual pada string domain. (Nadagoudar & Ramakrishna, 2024) juga mengevaluasi model *deep learning* berbasis RNN, LSTM, dan GRU untuk deteksi serta klasifikasi domain DGA.

Penelitian ini menggunakan *Random Forest* dengan fitur yang dihitung secara eksplisit sehingga perubahan kinerja pada *Unseen Family Split* dapat diamati berdasarkan karakteristik string yang digunakan model, bukan melalui representasi sekuens yang dipelajari secara otomatis.

Perbandingan Kinerja Kedua Skenario

Perbandingan pada Tabel 5 menunjukkan bahwa seluruh metrik mengalami penurunan pada *Unseen Family Split*. *Accuracy* berubah dari 91,83% menjadi 85,55%, *precision* dari 92,14% menjadi 91,47%, *recall* dari 91,45% menjadi 78,41%, *F1-score* dari 91,79% menjadi 84,44%, dan ROC-AUC dari 0,9653 menjadi 0,9066.

Perubahan terbesar terdapat pada *recall*, yaitu sebesar 13,04 poin persentase, sedangkan perubahan *precision* hanya sebesar 0,67 poin persentase. Pola tersebut menunjukkan bahwa perubahan kondisi pengujian lebih terlihat pada kemampuan model dalam menemukan domain DGA dibandingkan ketepatan ketika memberikan prediksi DGA.

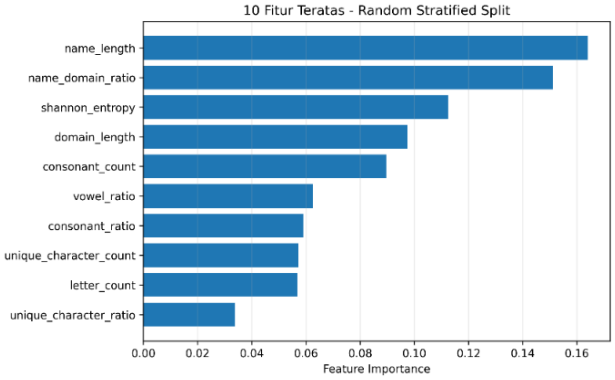
Perbedaan kondisi eksperimen perlu dipertimbangkan ketika membandingkan hasil dengan penelitian lain. (Pelayo-Benedet, Rodríguez, & Gañán, 2025) menekankan pentingnya konsistensi kondisi eksperimen dalam evaluasi deteksi AGD melalui pengembangan kerangka RAMPAGE untuk mendukung reproduktibilitas eksperimen.

Pendekatan yang berbeda juga menghasilkan karakteristik kinerja yang berbeda. (Xie et al., 2024) menggunakan *multi-stage feature fusion* dan memperoleh *accuracy* rata-rata 93,26% pada 33 *DGA family*, sedangkan (Jiang et al., 2025) menggunakan kombinasi *gated convolution* dan LSTM. Oleh karena itu, nilai metrik penelitian ini tidak dibandingkan secara langsung sebagai ukuran superioritas, tetapi digunakan untuk melihat perubahan kinerja model pada dua kondisi pembagian data.

(Pelayo-Benedet et al., 2025) dalam penelitian yang berbeda dari RAMPAGE juga menunjukkan bahwa LLM dapat mendeteksi AGD, tetapi kinerjanya dapat menurun ketika menghadapi trafik DNS dunia nyata dan domain *benign* yang memiliki struktur mencurigakan. Temuan tersebut menunjukkan bahwa perbedaan karakteristik data pengujian merupakan aspek yang perlu diperhatikan dalam interpretasi kinerja sistem deteksi domain.

Analisis Feature Importance

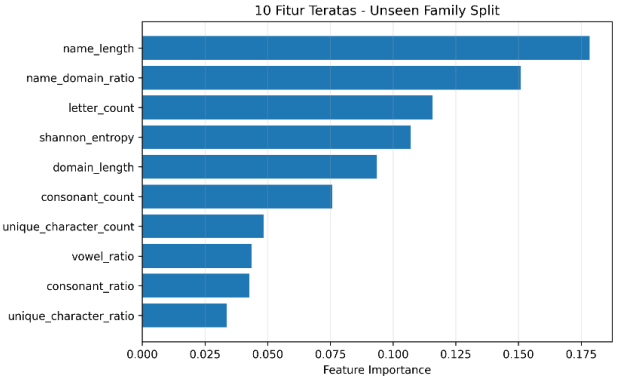
Analisis *feature importance* dilakukan untuk melihat kepentingan relatif dari 21 fitur yang digunakan oleh *Random Forest*. Sepuluh fitur dengan nilai *feature importance* tertinggi pada *Random Stratified Split* ditampilkan pada Gambar 4.



Gambar 4. Sepuluh fitur dengan feature importance tertinggi pada Random Stratified Split.

Berdasarkan Gambar 4, name length memiliki nilai *feature importance* tertinggi, diikuti oleh name-domain ratio dan Shannon entropy. Fitur lain yang termasuk dalam sepuluh nilai tertinggi adalah *domain length*, *consonant count*, *vowel ratio*, *consonant ratio*, *unique character count*, *letter count*, dan *unique character ratio*.

Analisis selanjutnya dilakukan pada *Unseen Family Split* untuk melihat perubahan kepentingan relatif fitur. Hasilnya ditampilkan pada Gambar 5.



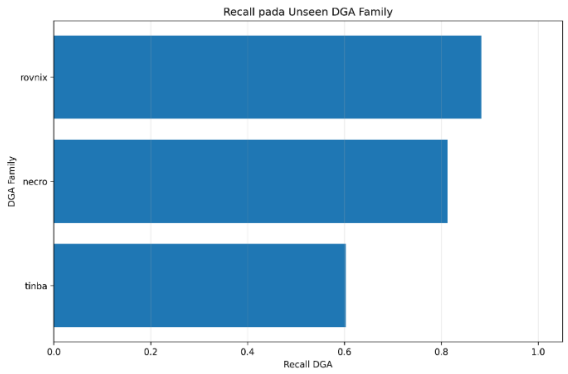
Gambar 5. Sepuluh fitur dengan feature importance tertinggi pada Unseen Family Split.

Pada Gambar 5, name length dan name-domain ratio tetap berada pada posisi teratas. Beberapa fitur lain mengalami perubahan urutan dibandingkan *Random Stratified Split*. Perubahan tersebut menunjukkan bahwa nilai kepentingan relatif fitur dapat berubah ketika komposisi *DGA family* pada data pengujian berbeda. Hasil tersebut menunjukkan bahwa fitur berbasis panjang dan komposisi string menjadi bagian penting dalam keputusan model. Namun, *feature importance* hanya menunjukkan kepentingan relatif fitur dalam model dan tidak digunakan untuk menyatakan hubungan kausal antara suatu fitur dan kelas DGA.

Temuan ini juga dapat dilihat dalam konteks penelitian (Fan et al., 2024), yang menunjukkan bahwa perbedaan karakteristik dan jumlah data antar-*DGA family* dapat memengaruhi kemampuan klasifikasi pada masing-masing kategori. Dengan demikian, perubahan urutan *feature importance* pada penelitian ini dapat dibaca sebagai perubahan kontribusi relatif fitur terhadap komposisi data yang digunakan pada masing-masing skenario.

Analisis Recall pada Unseen DGA Family

Evaluasi agregat menunjukkan adanya penurunan *recall* pada *Unseen Family Split*. Untuk melihat variasi hasil pada masing-masing *family*, dilakukan analisis tambahan dengan menghitung *recall* untuk setiap *DGA family* yang ditempatkan secara eksklusif pada data pengujian. Hasilnya ditampilkan pada Gambar 6.



Gambar 6. Recall pada DGA family yang tidak digunakan selama pelatihan.

Berdasarkan Gambar 6, nilai *recall* berbeda antar-*DGA family* meskipun model dan fitur yang digunakan sama. Perbedaan tersebut menunjukkan bahwa model menghasilkan tingkat deteksi yang berbeda ketika menghadapi pola domain dari *family* yang tidak tersedia pada data pelatihan.

Perbedaan tersebut juga perlu dipertimbangkan berdasarkan jumlah data pada masing-masing *family*. *DGA family* dengan jumlah sampel pengujian yang kecil dapat menghasilkan nilai *recall* yang berubah secara besar hanya akibat beberapa prediksi. Oleh karena itu, nilai *recall* per-*family* perlu diinterpretasikan bersama ukuran sampelnya dan tidak digunakan secara terpisah untuk menyimpulkan kemampuan model terhadap seluruh *DGA family*.

(Fan et al., 2024) menunjukkan bahwa ketidakseimbangan jumlah sampel antar-*DGA family* merupakan salah satu permasalahan dalam klasifikasi DGA karena kategori dengan jumlah sampel kecil dapat memperoleh kinerja yang berbeda dari kategori dengan jumlah sampel besar. Kondisi tersebut relevan dengan evaluasi *Unseen Family Split* karena jumlah domain pada setiap *family* pengujian tidak seragam.

Hasil *recall* per-*family* melengkapi hasil metrik agregat pada Tabel 5. Jika Tabel 5 menunjukkan perubahan kinerja model secara keseluruhan, Gambar 7 menunjukkan bahwa perubahan tersebut tidak terjadi secara seragam pada seluruh *DGA family*. Dengan demikian, evaluasi berdasarkan *family* memberikan informasi tambahan mengenai variasi kemampuan generalisasi model terhadap pola domain yang tidak diamati selama pelatihan.

5. KESIMPULAN DAN SARAN

Penelitian ini mengklasifikasikan domain DGA dan *legitimate* menggunakan *Random Forest* berdasarkan 21 fitur karakteristik string pada 20.000 domain, yang terdiri atas 10.000 domain DGA dari 52 *DGA family* dan 10.000 domain *legitimate*. Pada *Random Stratified Split*, model memperoleh *accuracy* 91,83%, *precision* 92,14%, *recall* 91,45%, *F1-score* 91,79%, dan ROC-AUC 0,9653, sedangkan *Unseen Family Split* dengan delapan *DGA family* yang tidak terdapat pada data pelatihan menghasilkan *accuracy* 85,55%, *precision* 91,47%, *recall* 78,41%, *F1-score* 84,44%, dan ROC-AUC 0,9066. Perbandingan kedua skenario menunjukkan penurunan *accuracy* sebesar 6,28 poin persentase dan *recall* sebesar 13,04 poin persentase pada pengujian terhadap *DGA family* yang tidak diamati selama pelatihan. *Feature importance* menunjukkan *name length* dan rasio nama domain terhadap panjang domain sebagai dua fitur dengan nilai kepentingan relatif tertinggi. Hasil tersebut menunjukkan bahwa *Unseen Family Split* memberikan informasi tambahan mengenai kinerja model terhadap *DGA family* di luar data pelatihan. Penelitian selanjutnya dapat mempertimbangkan penambahan variasi *DGA family*, sumber dataset yang berbeda, dan skenario pengujian berdasarkan perubahan waktu untuk mengevaluasi generalisasi model pada kondisi yang lebih beragam.

DAFTAR REFERENSI

Biros, H., & Kantor, M. (2025). Enhancing DGA detection with machine learning algorithms. *Journal of Telecommunications and Information Technology*, 31–44. https://doi.org/10.26636/jtit.2025.FITCE2024.2033

Chen, S., Lang, B., Chen, Y., & Xie, C. (2023). Detection of algorithmically generated malicious domain names with feature fusion of meaningful word segmentation and N-gram sequences. *Applied Sciences*, *13*(7). https://doi.org/10.3390/app13074406

Divya, T., Amritha, P. P., & Viswanathan, S. (2022). A model to detect domain names generated by DGA malware. *Procedia Computer Science*, *215*, 403–412. https://doi.org/10.1016/j.procs.2022.12.042

Fan, B., Ma, H., Liu, Y., Yuan, X., & Ke, W. (2024). KDTM: Multi-stage knowledge distillation transfer model for long-tailed DGA detection. *Mathematics*, *12*(5). https://doi.org/10.3390/math12050626

Jeremiah, D., Rafiq, H., Ta, V. T., Usman, M., Raza, M., & Awais, M. (2025). NIOM-DGA: Nature-inspired optimised ML-based model for DGA detection. *Computers & Security, 157*. https://doi.org/10.1016/j.cose.2025.104561

Jiang, K., Wu, S., Huang, R., & Deng, Z. (2025). DGA domain name detection model based on gated convolution and LSTM. *KSII Transactions on Internet and Information Systems*, *19*(3), 987–1006. https://doi.org/10.3837/tiis.2025.03.015

Lee, H., Do Yoo, J., Jeong, S., & Kim, H. K. (2024). Detecting domain names generated by DGAs with low false positives in Chinese domain names. *IEEE Access*, *12*, 123716–123730. https://doi.org/10.1109/ACCESS.2024.3454242

Liew, S. R. C., & Law, N. F. (2023). Use of subword tokenization for domain generation algorithm classification. *Cybersecurity*, *6*(1). https://doi.org/10.1186/s42400-023-00183-8

Nadagoudar, R. B., & Ramakrishna, M. (2024). DGA domain name detection and classification using deep learning models. *International Journal of Advanced Computer Science and Applications*, *15*. https://doi.org/10.14569/IJACSA.2024.0150730

Nie, Y., Liu, S., Qian, C., Deng, C., Li, X., Wang, Z., & Kuang, X. (2023). Multimodel collaboration to combat malicious domain fluxing. *Electronics*, *12*(19). https://doi.org/10.3390/electronics12194121

Pelayo-Benedet, T., Rodríguez, R. J., & Gañán, C. H. (2025). The machines are watching: Exploring the potential of large language models for detecting algorithmically generated domains. *Journal of Information Security and Applications*, *93*. https://doi.org/10.1016/j.jisa.2025.104176

Peng, X., He, J., Ni, L., & Yang, G. (2026). Beyond pattern matching: A cognitive-driven framework for DGA detection via dual-perspective anomaly perception. *Electronics*, *15*(9). https://doi.org/10.3390/electronics15091934

Selvaraj, S., & Panjanathan, R. (2024). WordDGA: Hybrid knowledge-based word-level domain names against DGA classifiers and adversarial DGAs. *Informatics*, *11*(4). https://doi.org/10.3390/informatics11040092

Sun, X., & Liu, Z. (2023). Domain generation algorithms detection with feature extraction and Domain Center construction. *PLOS ONE*, *18*(1). https://doi.org/10.1371/journal.pone.0279866

Suryotrisongko, H., Musashi, Y., Tsuneda, A., & Sugitani, K. (2022). Robust botnet DGA detection: Blending XAI and OSINT for cyber threat intelligence sharing. *IEEE Access*, *10*, 34613–34624. https://doi.org/10.1109/ACCESS.2022.3162588

Tang, J., Guan, Y., Zhao, S., Wang, H., & Chen, Y. (2024). DGA domain detection based on Transformer and rapid selective kernel network. *Electronics*, *13*(24). https://doi.org/10.3390/electronics13244982

Vranken, H., & Alizadeh, H. (2022). Detection of DGA-generated domain names with TF-IDF. *Electronics*, *11*(3). https://doi.org/10.3390/electronics11030414

Wang, Z., Guo, Y., & Montgomery, D. (2022). Machine learning-based algorithmically generated domain detection. *Computers and Electrical Engineering*, *100*. https://doi.org/10.1016/j.compeleceng.2022.107841

Xie, M., He, R., & He, A. (2024). Deep learning DGA malicious domain name detection based on multi-stage feature fusion. *Applied and Computational Engineering*, *64*(1), 1–8. https://doi.org/10.54254/2755-2721/64/20241334

Yang, C., Lu, T., Yan, S., Zhang, J., & Yu, X. (2022). N-Trans: Parallel detection algorithm for DGA domain names. *Future Internet*, *14*(7). https://doi.org/10.3390/fi14070209