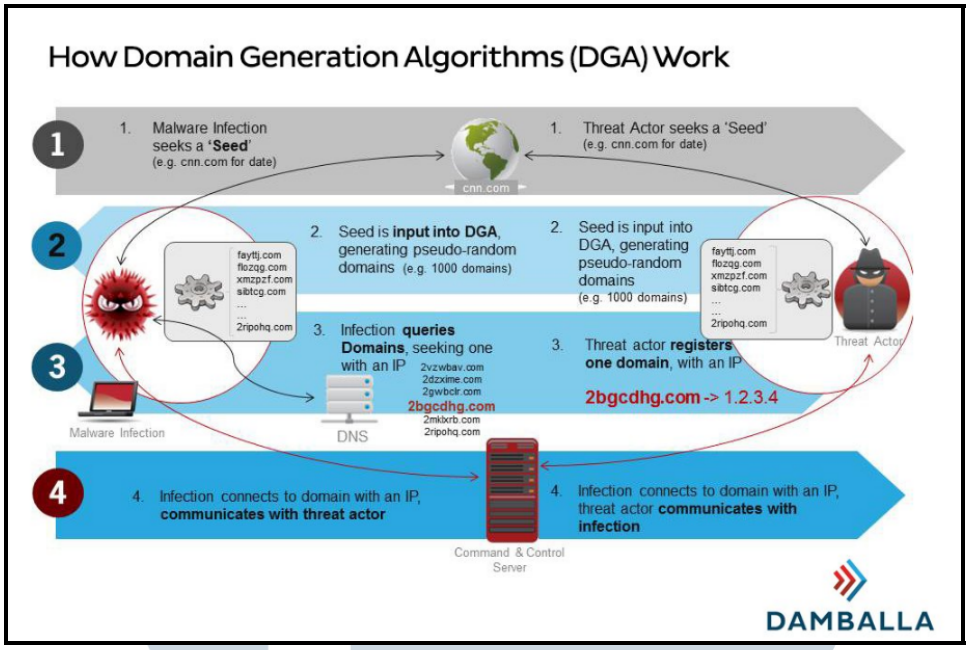
BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Kemunculan serangan siber yang semakin canggih, seperti *botnet* dan *ransomware*, menjadi tantangan utama dalam keamanan informasi saat ini [3]. Serangan-serangan tersebut umumnya mengandalkan infrastruktur *Command-and-Control* (C&C) untuk mengendalikan aktivitas berbahaya pada sistem yang telah terinfeksi secara jarak jauh. Untuk menghindari pendeteksian dan penindakan oleh pihak *defender*, penyerang terus mengembangkan teknik komunikasi terselubung. Salah satu teknik yang banyak digunakan adalah *Domain Generation Algorithm* (DGA) [1]. Teknik DGA memungkinkan *malware* menghasilkan sejumlah besar nama domain secara acak atau *pseudo-random* untuk mencari koneksi ke server C&C miliknya [1, 4]. Karena nama-nama domain ini berubah-ubah dan tampak acak, upaya pemblokiran melalui mekanisme keamanan tradisional (seperti *blacklist* domain) menjadi sangat sulit [1]. Akibatnya, *malware* yang menggunakan DGA dapat tetap beroperasi meskipun beberapa nama domain berhasil diblokir, karena tersedia domain lain sebagai kanal cadangan untuk melanjutkan komunikasi dengan server C&C [5, 4].

Gambar 1.1 menunjukkan ilustrasi alur kerja *Domain Generation Algorithm* (DGA) pada *malware*. Secara ringkas, *malware* membangkitkan banyak kandidat nama domain secara periodik, misalnya berdasarkan waktu atau benih tertentu, lalu melakukan resolusi DNS terhadap daftar domain tersebut. Penyerang cukup mendaftarkan sebagian kecil domain yang sesuai dengan keluaran DGA pada hari tersebut, sehingga korban dapat kembali terhubung ke server *Command-and-Control* (C&C) tanpa menggunakan domain statis yang mudah diblokir. Oleh karena domain yang digunakan berubah-ubah dan jumlahnya besar, pemblokiran konvensional berbasis daftar hitam (*blacklist*) menjadi kurang efektif untuk kasus DGA.



Gambar 1.1. Cara kerja *Domain Generation Algorithm*. Diadaptasi dari [1].

Banyak keluarga *malware* mutakhir termasuk *botnet*, *trojan*, hingga *ransomware* diketahui mengintegrasikan DGA untuk menyembunyikan jejak komunikasinya dengan server C&C [1, 4]. Dengan memanfaatkan DGA, penyerang hanya perlu mendaftarkan sebagian kecil domain yang valid untuk infrastruktur C&C, sementara *malware* akan menghasilkan dan mencoba banyak kandidat domain lain [4]. Hal ini menimbulkan *noise* berupa banyak kueri DNS yang gagal atau *unresolved* misalnya *NXDOMAIN* serta pola kueri yang tidak lazim [1]. Teknik ini membuat pertahanan konvensional seperti pemblokiran domain secara manual, *sinkhole* DNS, atau aturan *Intrusion Detection System* (IDS) menjadi kurang efektif bila dilakukan secara statis atau manual terhadap daftar domain, karena skala domain yang dapat dihasilkan sangat besar dan berubah-ubah [1]. Dengan kata lain, sangat tidak praktis bagi *defender* untuk memprediksi dan memblokir seluruh kemungkinan domain yang dihasilkan secara dinamis oleh DGA.

Untuk menghadapi ancaman yang adaptif ini, para peneliti mengembangkan metode deteksi *malicious domain* berbasis *machine learning* yang menganalisis pola karakter pada nama domain guna membedakan domain hasil DGA dari domain normal [6, 4]. Berbeda dengan daftar *blacklist* statis, pendekatan *machine learning* mampu mengenali ciri-ciri khas dari domain jahat meskipun nama domain tersebut belum pernah muncul sebelumnya. Ciri-ciri (fitur) yang umum digunakan meliputi panjang nama domain, tingkat keragaman karakter (entropi), frekuensi

kemunculan *n-gram*, serta fitur leksikal atau statistik lain yang membedakan domain acak dengan domain yang lazim digunakan manusia [6, 4]. Dengan teknik ini, sejumlah sistem deteksi dilaporkan mencapai tingkat akurasi yang tinggi. Sebagai contoh, MaldomDetector mampu mendeteksi domain hasil DGA dengan akurasi sekitar 98% menggunakan pengukuran keacakan karakter domain dan fitur teks sederhana [5]. Penelitian terkini juga melaporkan model yang dioptimasi mampu mempertahankan akurasi tinggi saat diuji pada skala sangat besar (hingga puluhan juta domain) [7]. Hal ini menegaskan bahwa pendekatan *machine learning* memiliki potensi besar dalam mengidentifikasi domain-domain mencurigakan secara otomatis sebelum sempat digunakan oleh *malware* untuk komunikasi C&C.

Walaupun berbagai pendekatan deteksi DGA telah berhasil mencapai kinerja yang menjanjikan, setiap metode memiliki kelemahan dan *trade-off* tersendiri. Metode berbasis rekayasa fitur (*feature engineering*) dengan algoritma klasik seperti *Random Forest* terbukti efektif dan relatif mudah diinterpretasi, namun perancangan fitur yang tepat tidaklah mudah dan tetap ada risiko degradasi saat berhadapan dengan variasi DGA yang berbeda [6, 4]. Di sisi lain, metode *deep learning* dapat mempelajari pola langsung dari data mentah (karakter domain) dan sering menghasilkan performa tinggi, termasuk untuk variasi DGA yang lebih kompleks [4]. Akan tetapi, model *deep learning* murni sering menghadapi tantangan interpretabilitas (cenderung *black box*) dan isu efektivitas pada DGA berbasis kata/*pseudo-words* sehingga memotivasi eksplorasi representasi/tokenisasi yang lebih sesuai [6, 4]. Kondisi ini menunjukkan perlunya evaluasi komparatif untuk menimbang model mana yang paling optimal.

Berdasarkan paparan di atas, terdapat kebutuhan mendesak untuk mengevaluasi efektivitas model deteksi dalam skenario yang lebih realistis menggunakan dataset gabungan (*unified dataset*) yang mencakup berbagai varian DGA. Penelitian ini bertujuan melakukan analisis komparatif mendalam antara pendekatan *machine learning* berbasis fitur (*Random Forest*) dan *deep learning* (BiLSTM), serta mengusulkan pendekatan *Hybrid* yang menggabungkan keunggulan ekstraksi fitur otomatis sekuensial dengan fitur statistik rekayasa (*handcrafted features*). Tidak hanya berfokus pada metrik akurasi semata, penelitian ini juga menerapkan pendekatan *Explainable AI* (XAI) untuk mengatasi masalah *black box*, memberikan transparansi mengenai bagaimana model membedakan domain DGA yang kompleks dari domain yang sah.

Berdasarkan komparasi penelitian terdahulu yang dirangkum pada Tabel 1.1,

tren penelitian deteksi *Domain Generation Algorithm* (DGA) dalam periode 2020–2025 didominasi oleh pendekatan *Deep Learning* yang kompleks serta algoritma klasik berbasis fitur. Meskipun berbagai pendekatan tersebut menawarkan arsitektur yang canggih, terdapat kesenjangan signifikan dari aspek transparansi model. Mayoritas studi beroperasi sebagai *black-box* dan belum mengintegrasikan komponen *Explainable AI* (XAI), kecuali pada penelitian Jeremiah et al. [7]. Ketiadaan fitur transparansi ini menyebabkan keputusan model dalam mengklasifikasikan domain sebagai DGA atau *legit* sulit diverifikasi oleh analis keamanan.

Indikasi DGA dapat dianalisis dari dua sudut pandang, yaitu berbasis telemetri DNS (misalnya lonjakan *NXDOMAIN* atau pola kueri tidak lazim) dan berbasis string nama domain (*domain-only*). Pendekatan telemetri umumnya membutuhkan log DNS yang lengkap dan stabil, sehingga dapat dipengaruhi konfigurasi resolver, caching, serta pola penggunaan jaringan. Sebaliknya, pendekatan domain-only memanfaatkan ciri leksikal domain seperti tingkat keacakan, distribusi karakter, dan pola *n-gram*, sehingga lebih mudah diterapkan ketika data trafik tidak tersedia. Karena itu, penelitian ini memfokuskan deteksi pada *Second Level Domain* (SLD).

Selain isu transparansi, terdapat pula inkonsistensi dalam metodologi pengelolaan data. Banyak penelitian terdahulu menggunakan dataset dengan jumlah terbatas, tidak seimbang (*imbalanced*), atau tidak menerapkan validasi yang ketat seperti *K-Fold Cross-Validation*. Hal ini membatasi keandalan model saat menghadapi variasi serangan DGA yang masif di dunia nyata. Berangkat dari kesenjangan metodologis tersebut, penelitian ini mengajukan pendekatan deteksi yang tidak hanya menekankan pada efisiensi seleksi fitur dan penanganan ketidakseimbangan data, tetapi juga menjamin interpretabilitas keputusan model melalui integrasi metode XAI.

Berdasarkan Tabel 1.1, terdapat beberapa *research gap* yang menjadi landasan penelitian ini. Pertama, sebagian penelitian terdahulu berfokus pada capaian akurasi, tetapi belum menekankan aspek interpretabilitas (*explainable AI*/XAI) untuk menjelaskan alasan prediksi model. Kedua, skema validasi eksperimen pada berbagai studi masih bervariasi, sehingga perbandingan kinerja antarmetode menjadi kurang sebanding apabila evaluasi tidak dilakukan secara konsisten. Ketiga, pada skenario data berskala besar, kebutuhan seleksi fitur dan efisiensi komputasi menjadi faktor penting agar solusi lebih realistis untuk diadopsi. Oleh sebab itu, penelitian ini menyusun eksperimen komparatif Random

Tabel 1.1. Tinjauan metodologi penelitian terdahulu

Penelitian	Data & Validasi	Fitur	Algoritma	XAI?
Yang et al. 2020	500k (Imbalanced), 90/10 Split	Embedding	CNN + BiLSTM	Tidak
Tang et al. 2024	110k (Balanced), 80/20 Split	Embedding	Transformer	Tidak
Jeremiah et al. 2025	16 Juta (Balanced), 80/20 Split	78 Fitur Handcrafted	Random Forest	✓
Sun & Liu 2023	4.3 Juta (Balanced), Tanpa CV	35 Fitur	BiLSTM + CNN	Tidak
Namgung et al. 2021	4 Juta (Imbalanced), 80/10/10	Char Embedding	CNN, BiLSTM	Tidak
Kuna et al. 2025	280k (Balanced), 10-Fold	Domain/Host	Random Forest	Tidak
Vu & Hoang 2025	500k (Balanced), Sampling	TF-IDF	XGBoost + BERT	Tidak
Hwang et al. 2020	45k (Imbalanced), 90/10	10 Fitur	TextCNN	Tidak
Leyva et al. 2024	136k (Balanced), 5-Split	Raw Domain	LLM (Llama3)	Tidak
Nadagoudar 2024	20k (Balanced), 5-Percobaan	Embedding	GRU	Tidak
Gregorio et al. 2024	12k (Balanced), -	Char Embedding	CNN	Tidak
Chen et al. 2023	190k (Imbalanced), Random	N-gram	XGBoost	Tidak
Highnam et al. 2021	68k (Imbalanced), Predefined	Deep Features	CNN + LSTM	Tidak
Gadre et al. 2024	50k (Balanced), 80/20	Embedding	Transformer	Tidak

Forest dan BiLSTM pada *unified dataset*, menerapkan evaluasi yang konsisten, serta menambahkan analisis interpretabilitas menggunakan SHAP. Ringkasan keterkaitan *research gap* tersebut dengan hasil penelitian akan dibahas kembali pada Bab 4 (lihat Tabel 4.7).

Ringkasan pada Bab 4 tidak hanya memetakan *research gap*, tetapi juga menunjukkan bagian mana yang telah dijawab melalui hasil eksperimen dan analisis interpretabilitas.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini dapat dirincikan ke dalam beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimana perbandingan kinerja algoritme berbasis fitur (*Random Forest*) dan *deep learning* (BiLSTM) dalam mengklasifikasikan nama domain pada *unified dataset* yang memiliki variasi karakteristik DGA yang beragam?
2. Apakah penggabungan fitur statistik rekayasa (*handcrafted features*) dengan pembelajaran fitur sekuensial otomatis pada arsitektur *Hybrid* BiLSTM mampu meningkatkan performa deteksi dibandingkan dengan model BiLSTM murni (*Pure* BiLSTM)?
3. Bagaimana mekanisme pengambilan keputusan model dalam mendeteksi DGA dapat dijelaskan secara transparan menggunakan pendekatan *Explainable AI* (XAI) untuk mengidentifikasi fitur dominan yang membedakan domain DGA dan *legit*?

1.3 Batasan Permasalahan

Agar fokus penelitian lebih terarah dan konsisten dengan metodologi yang diterapkan, beberapa batasan masalah ditetapkan sebagai berikut:

1. Penelitian ini dibatasi pada pendeteksian domain hasil DGA sebagai klasifikasi biner (*binary classification*) untuk menentukan apakah suatu nama domain tergolong DGA (*malicious*) atau *legit* (*benign*). Penelitian tidak mencakup klasifikasi multi-kelas untuk mengidentifikasi famili *malware* secara spesifik.

2. Dataset yang digunakan merupakan unifikasi data sekunder dari berbagai repositori publik dan dataset akademik. Sampel domain DGA bersumber dari *UMUDGA*, *Data Driven Security*, *Kaggle*, dan dataset *ExtraHop*. Sementara itu, sampel domain *legit* bersumber dari daftar *Tranco Top-1M* dan arsip *Alexa Top-1M*. Penggunaan berbagai sumber ini bertujuan merepresentasikan variasi karakteristik domain dalam skenario *unified dataset*.

3. Ruang lingkup analisis karakter domain dibatasi pada himpunan karakter standar (*ASCII*/Latin) yang umum digunakan. Penelitian ini tidak mencakup analisis terhadap *Internationalized Domain Names* (IDN) yang menggunakan aksara non-Latin, seperti Mandarin, Cyrillic, dan Arab. Selain itu, fitur berbasis kamus (*dictionary-based features*) secara spesifik hanya menggunakan kosakata bahasa Inggris (*English wordlist*).

4. Fokus analisis dilakukan pada bagian *Second Level Domain* (SLD). Informasi *Top Level Domain* (TLD) dan *subdomain* dieliminasi pada tahap pra-pemrosesan untuk memfokuskan deteksi pada pola string utama dan menghindari bias ekstensi domain. Pendekatan ini bersifat statis, meliputi ekstraksi fitur leksikal dan *n-gram*, tanpa melibatkan analisis jaringan dinamis seperti inspeksi paket (*deep packet inspection*) atau kueri DNS waktu nyata (*real-time*).

5. Penelitian ini membandingkan dua pendekatan algoritme, yaitu *Random Forest* (representasi *ensemble learning* berbasis fitur) dan *Bidirectional LSTM* (BiLSTM) (representasi *deep learning* berbasis sekuensial). Algoritme lain seperti *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), serta arsitektur lain (misalnya CNN, GRU, atau *transformer*) tidak dibahas.

6. Evaluasi kinerja model diukur menggunakan metrik standar, yaitu akurasi, presisi, *recall*, F1-Score, dan ROC-AUC. Interpretabilitas dievaluasi menggunakan metode SHAP (*SHapley Additive exPlanations*), tanpa membahas latensi inferensi atau konsumsi memori secara mendalam.

7. Penelitian ini bersifat eksperimental dan seluruh proses pengembangan serta evaluasi model dilakukan secara *offline* menggunakan dataset publik. Penelitian ini tidak mencakup implementasi sistem pada *enterprise network* secara langsung maupun pengujian terhadap serangan aktif (*live attack*).

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, tujuan utama dari penelitian ini adalah:

1. Mengukur dan membandingkan kinerja model *Random Forest* dan *BiLSTM* dalam mendeteksi domain DGA pada dataset gabungan (*unified dataset*) untuk mengetahui algoritme mana yang paling efektif menangani variasi karakteristik DGA yang heterogen.

2. Mengevaluasi dampak pendekatan *Hybrid* dengan menganalisis apakah penggabungan fitur statistik rekayasa (*handcrafted features*) memberikan kontribusi signifikan atau justru redundan dibandingkan dengan arsitektur *Pure* BiLSTM yang mengandalkan ekstraksi fitur sekuensial otomatis.

3. Menganalisis perilaku pengambilan keputusan model secara transparan menggunakan metode *Explainable AI* (SHAP) guna memvisualisasikan fitur dominan dan pola karakter yang menjadi penentu dalam klasifikasi domain DGA maupun *legit*.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan memberikan kontribusi teoretis dan praktis sebagai berikut.

Manfaat Teoretis

1. Memberikan analisis komparatif mengenai efektivitas *machine learning* berbasis rekayasa fitur (Random Forest) dibandingkan *deep learning*

(BiLSTM) pada *unified dataset*, sehingga memperjelas *trade-off* antara interpretabilitas dan performa.

2. Memberikan wawasan empiris mengenai hubungan antara rekayasa fitur (*feature engineering*) dan pembelajaran fitur otomatis (*feature learning*), termasuk kondisi ketika penambahan fitur statistik bersifat signifikan atau mulai mengalami *diminishing returns*.

3. Memperkaya literatur keamanan siber terkait penerapan interpretabilitas model menggunakan SHAP untuk meningkatkan transparansi model yang cenderung *black-box*.

Manfaat Praktis

1. Menyediakan rekomendasi pemilihan model berdasarkan kebutuhan operasional, misalnya prioritas menekan *false positive* pada domain *legit* atau menekan *false negative* pada domain DGA.

2. Memberikan wawasan mengenai fitur dominan hasil seleksi fitur dan analisis SHAP, yang dapat membantu analis keamanan merancang aturan deteksi (*rule*) yang lebih efisien.

3. Menghasilkan model deteksi yang diuji stabilitasnya melalui validasi silang (*cross-validation*), sehingga lebih siap menghadapi variasi ancaman DGA pada lingkungan nyata.

1.6 Sistematika Penulisan

Laporan penelitian ini disusun dalam lima bab. Bab 1 memuat pendahuluan. Bab 2 membahas landasan teori dan penelitian terkait. Bab 3 menjelaskan metodologi penelitian dan rancangan eksperimen. Bab 4 memaparkan hasil, analisis, dan pembahasan. Bab 5 berisi simpulan dan saran.