BAB I
PENDAHULUAN

1.1 Latar Belakang

Serangan *Distributed Denial of Service* (DDoS) telah menjadi ancaman serius dalam keamanan jaringan, dengan trend serangan yang terus meningkat dalam beberapa tahun terakhir (Kalutharage dkk., 2023). Serangan ini bekerja dengan membanjiri jaringan, server, atau aplikasi menggunakan *botnet* berupa kumpulan perangkat yang dikendalikan oleh penyerang untuk menciptakan arus lalu lintas yang masif, sehingga mengancam aspek *availability* dalam keamanan informasi (Jithu dkk., 2021). Dampak paling krusial dari serangan DDoS adalah terganggunya operasional layanan, hilangnya akses pengguna, hingga kerugian finansial besar. Contoh ekstrem dari serangan ini adalah serangan DDoS terhadap penyedia layanan DNS terkemuka, Dyn pada tahun 2016 yang menyebabkan layanan besar seperti Twitter, Spotify, Netflix, PayPal, Visa, Amazon dan Reddit mengalami *downtime* masif, membuktikan bagaimana serangan ini dapat melumpuhkan infrastruktur digital dalam skala global (Yeh dkk., 2020).

Pada tahun 2024, berdasarkan laporan resmi dari situs *Cloudflare*, oleh Yoachimik & Pacheco, (2025) bahwa *Cloudflare* telah memblokir sekitar 21,3 juta serangan DDoS, meningkat 53% dibandingkan tahun sebelumnya, dengan rata-rata 4.870 serangan per jam. Lebih dari 420 serangan bersifat hipervolumetrik, melampaui 1 miliar paket per detik (pps) dan 1 Tbps, sementara jumlah serangan di atas 1 Tbps meningkat drastis sebesar 1.885% dari kuartal sebelumnya. Secara spesifik, serangan DDoS pada *Layer* 3/*Layer* 4 berjumlah 49% atau 3,4 juta, dengan vektor utama yaitu SYN Flood 38%, DNS Flood 16%, dan UDP Flood 14%. Sebanyak 51% atau 3,5 juta lainnya adalah serangan DDoS HTTP, di mana 11% berpura-pura sebagai *browser* yang sah, 10% memiliki atribut HTTP mencurigakan atau tidak biasa, dan 8% terdiri dari banjir HTTP generik, serangan pembobolan *cache volumetrik*, serta serangan yang menargetkan titik akhir *login*. Menariknya, Sekitar 14% permintaan HTTP dengan metode HEAD dan 8% dengan metode DELETE yang merupakan bagian dari serangan DDoS, meskipun metode ini hampir tidak ada dalam permintaan HTTP yang sah.

Dalam menghadapi serangan DDoS, salah satu upaya yang dapat dilakukan adalah dengan menerapkan *firewall* dan *Access Control List* (ACL). Keduanya bekerja dengan menyaring lalu lintas jaringan berdasarkan aturan tertentu, seperti mengizinkan atau memblokir alamat IP, nomor port, atau jenis protokol tertentu (Quadir dkk., 2020). Namun, pendekatan ini memiliki keterbatasan, terutama dalam menangani serangan DDoS yang bersifat terdistribusi dan berasal dari berbagai alamat IP berbeda (*botnet*) (Sassani dkk., 2022). Oleh karena itu, diperlukan pendekatan yang lebih adaptif dan efektif dalam mendeteksi serangan tersebut.

Salah satu pendekatan yang efektif adalah dengan penerapan *machine learning*. Teknologi ini memungkinkan sistem untuk belajar dari data tanpa diprogram secara eksplisit, sehingga mampu menganalisis pola lalu lintas jaringan secara dinamis dan membedakan antara lalu lintas normal dengan lalu lintas serangan (Najafimehr dkk., 2023). *Machine learning* unggul karena kemampuannya mengenali pola-pola kompleks yang seringkali tidak teridentifikasi oleh metode tradisional (Bhayo dkk., 2023). Berbagai algoritma telah digunakan untuk klasifikasi serangan DDoS, seperti *Support Vector Machine* (SVM), *k-Nearest Neighbors* (KNN), *Naive Bayes* (NB), dan *Random Forest* (RF) (Ismail dkk., 2022). Penelitian oleh Maslan dkk., (2020) menunjukkan bahwa algoritma *Random Forest* menghasilkan akurasi tertinggi sebesar 98,7% dalam mendeteksi serangan DDoS pada *dataset* baru dibandingkan algoritma klasifikasi lainnya.

Random Forest (RF) merupakan algoritma yang umum digunakan dalam klasifikasi serangan DDoS karena ketangguhannya terhadap *noise* serta kemampuannya beradaptasi dengan berbagai jenis dan karakteristik data (Baskaya & Samet, 2020). Algoritma ini bekerja dengan membangun *decision tree* (DT) dan menggabungkan hasilnya untuk memperoleh klasifikasi yang lebih akurat (Georganos dkk., 2021). Selain mampu menangani dataset besar dengan banyak fitur, *Randon Forest* juga efektif dalam mengurangi resiko *overfitting* (Oo & Thein, 2022). Penelitian oleh Fathima dkk., (2023) membandingkan performa *Random Forest*, *KNN*, dan *regresi logistik* menggunakan *dataset* CSE-CICIDS2018, CSE-CICIDS2017, dan CICDoS. Hasilnya, *random forest* mencapai akurasi tertinggi sebesar 97,6%, diikuti KNN (97%) dan *regresi logistik* (91,1%), menegaskan keunggulan *random forest* dalam menghadapi data berskala besar dan kompleks.

Meski demikian, untuk meningkatkan efisiensi dan akurasi model, diperlukan proses seleksi fitur. Penelitian oleh Awad & Fraihat, (2023) membuktikan bahwa penerapan *Recursive Feature Elimination* (*RFE*) dengan *Cross-Validation* dan *Decision Tree* sebagai *estimator* mampu menghasilkan akurasi klasifikasi sebesar 95,30%, hanya sedikit lebih rendah dibandingkan penggunaan seluruh fitur (95,56%). Temuan ini menunjukkan bahwa pengurangan jumlah fitur tidak selalu mengorbankan akurasi secara signifikan. Akan tetapi, meskipun efisiensi meningkat dari segi jumlah fitur, isu terkait efisiensi waktu pemrosesan (*computational cost*) belum menjadi fokus utama. Dengan demikian, masih dibutuhkan pendekatan seleksi fitur yang mampu meningkatkan efisiensi komputasi tanpa mengurangi performa model secara signifikan (Koul & Manvi, 2020).

Salah satu metode yang dapat dilakukan adalah dengan kombinasi seleksi fitur *Mutual Information* dan RFE. Penelitian oleh Koul & Manvi, (2020) berhasil mengurangi dimensi data ekspresi gen kanker sebesar 98,5% melalui kombinasi kedua metode ini dari ribuan gen menjadi hanya 316 gen serta menghasilkan akurasi tertinggi mencapai 99%. *Mutual Information* berperan dalam mengeliminasi fitur tidak relevan secara awal, sementara *RFE* mempertahankan fitur yang paling berpengaruh terhadap performa model. Hasilnya, waktu pemrosesan lebih cepat dan model prediksi menjadi lebih efisien dan akurat, menunjukkan keunggulan metode ini dalam diagnosis dini penyakit kanker.

Berdasarkan temuan penelitian sebelumnya, penelitian ini bertujuan untuk mengembangkan model klasifikasi biner serangan DDoS dari dataset CIC-DDoS2019 yang telah disederhanakan, sehingga penelitian ini memiliki keterbatasan dalam representasi data serangan DDoS yang lengkap. Penelitian ini berjudul **"IMPLEMENTASI ALGORITMA *RANDOM FOREST* DALAM KLASIFIKASI SERANGAN *DISTRIBUTED DENIAL OF SERVICE* MENGGUNAKAN KOMBINASI SELEKSI FITUR"** Evaluasi performa model dilakukan dengan metrik seperti *Accuracy*, *Precision*, *Recall, F1-Score*, ROC-AU, *Detection time,* dan *Processing time*. Pendekatan ini diharapkan dapat menghasilkan model yang tidak hanya akurat dalam klasifikasi, tetapi juga efisien dari sisi waktu dan komputasi melalui pengurangan dimensi fitur secara optimal.

1.2 Rumusan Masalah

1. Bagaimana pengaruh kombinasi *Mutual Information* dan *Recursive Feature Elimination Cross-Validation* dapat menghasilkan subset fitur yang optimal dan menyederhanakan proses komputasi?
2. Bagaimana performa algoritma *Random Forest* dalam mengklasifikasikan serangan DDoS setelah diterapkan kombinasi seleksi fitur *Mutual Information* dan *Recursive Feature Elimination Cross-Validation*?

1.3 Tujuan Penelitian

1. Menganalisis pengaruh kombinasi Mutual Information dan Recursive Feature Elimination Cross-Validation dalam menghasilkan subset fitur yang optimal dan menyederhanakan proses komputasi.
2. Mengevaluasi performa algoritma Random Forest dalam mengklasifikasikan serangan DDoS setelah penerapan kombinasi seleksi fitur Mutual Information dan Recursive Feature Elimination Cross-Validation.
3. Mengukur efisiensi model setelah penerapan seleksi fitur untuk melihat pengaruhnya terhadap kompleksitas komputasi, waktu pemrosesan dan waktu deteksi.

1.4 Manfaat Penelitian

1. Memberikan kontribusi pada literatur akademis di bidang keamanan siber, khususnya dalam penerapan *machine learning* untuk deteksi serangan DDoS.
2. Membantu pengembang sistem keamanan dalam memilih metode seleksi fitur terbaik untuk meningkatkan akurasi dan efisiensi model deteksi serangan DDoS
3. Mengurangi kompleksitas pemrosesan data melalui seleksi fitur yang optimal, sehingga dapat diterapkan pada infrastruktur teknologi informasi yang lebih luas.
4. Meningkatkan efisiensi dalam pengelolaan data besar yang sering muncul dalam analisis serangan DDoS dengan memanfaatkan seleksi fitur.

1.5 Batasan Penelitian

1. Penelitian ini hanya berfokus pada klasifikasi serangan DDoS dan normal secara umum, tanpa mengklasifikasikan jenis-jenis serangan DDoS secara spesifik.
2. Penelitian ini belum menguji kemampuan generalisasi model terhadap jenis serangan DDoS yang lebih baru dan kompleks yang tidak tercakup dalam dataset CIC-DDoS2019.
3. Penelitian ini menggunakan *dataset* CIC-DDoS2019 yang berukuran 96,1Mb, yang merupakan hasil penyederhanaan dari dataset asli berukuran 30GB. Dataset ini tidak mencakup seluruh jenis serangan dan traffic DDoS yang tersedia pada versi aslinya. Oleh karena itu, hasil klasifikasi tidak dapat digeneralisasi secara menyeluruh terhadap semua skenario serangan DDoS di dunia nyata.
4. Penelitian ini hanya berfokus pada klasifikasi serangan DDoS dan tidak mencakup serangan siber lainnya, seperti serangan brute force, phishing, atau serangan berbasis malware.
5. Seleksi fitur hanya dilakukan menggunakan metode Mutual Information dan *Recursive Feature Elimination Cross-Validation*, tanpa membandingkannya dengan metode seleksi fitur lainnya.

1.6 Struktur Organisasi Skripsi

Sistematika penulisan skripsi ini diantaranya sebagai berikut.

1. BAB I PENDAHULUAN

   Bab pendahuluan yang terdiri dari penjelasan latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, dan struktur organisasi skripsi.

2. BAB II TINJAUAN PUSTAKA

   Bab tinjauan pustaka memberikan pembahasan terkait topik serta permasalahan yang diangkat pada penelitian. Bab ini berisikan literature review dari penelitian-penelitian terhadulu yang berkaitan untuk memberikan gambaran *state-of-the-art* dari penelitian ini.

3. BAB III METODOLOGI PENELITIAN

   Bab metodologi penelitian membahas alur penelitian dari metode yang akan digunakan serta tahapan-tahapan dalam penelitian mulai dari awal hingga akhir.

4. BAB IV HASIL DAN PEMBAHASAN

   Bab hasil dan pembahasan menyampaikan hasil penelitian yang sesuai dengan perumusan masalah di awal bab pertama.

5. BAB V KESIMPULAN DAN SARAN

   Bab kesimpulan dan saran memberikan pemaparan terkait kesimpulan dari hasil penelitian yang telah dilakukan, implikasi, serta memberikan rekomendasi untuk penelitian mendatang.