BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi yang semakin pesat telah membawa perubahan signifikan dalam berbagai sektor kehidupan, termasuk ekonomi digital, efisiensi operasional organisasi, serta transformasi layanan publik. Pemanfaatan teknologi jaringan dan internet memungkinkan pertukaran data secara cepat dan masif, namun pada saat yang sama turut meningkatkan risiko terhadap keamanan sistem informasi. Kompleksitas infrastruktur digital yang terus berkembang menjadikan keamanan siber sebagai aspek krusial yang harus dikelola secara sistematis dan berkelanjutan (Abdurrasyid, Susanti, et al., 2024; Fadhlurrohman et al., 2024).

Berbagai penelitian menunjukkan bahwa peningkatan konektivitas dan digitalisasi yang tidak diimbangi dengan mekanisme keamanan yang memadai dapat membuka celah terhadap berbagai jenis serangan siber (Antony et al., 2024; Sudirman, 2024). Salah satu ancaman keamanan siber yang memiliki dampak signifikan adalah serangan botnet. Botnet merupakan kumpulan perangkat yang terinfeksi malware dan dikendalikan secara terpusat oleh pihak tertentu untuk melakukan aktivitas berbahaya, seperti pencurian data, penyebaran spam, hingga serangan Distributed Denial of Service (DDoS) (Mudassir et al., 2022; Tatarnikova et al., 2022). Skala dan sifat terdistribusi dari botnet menjadikannya sulit untuk dideteksi dan dikendalikan. Menurut laporan Cybersecurity Ventures, diperkirakan bahwa kerugian global akibat kejahatan siber akan mencapai lebih dari 10,5 triliun dolar Amerika Serikat per tahun pada 2025, dengan botnet tetap menjadi salah satu kontributor utama peningkatan kerugian tersebut (Najmul et al., 2025; Nasdaq, 2020; Vergara Cobos et al., n.d.). Menurut Chee et al. (2025) dan Georgoulias et al. (2023), kondisi ini menegaskan bahwa botnet merupakan faktor dominan dalam eskalasi ancaman siber global.

Deteksi botnet pada lalu lintas jaringan menjadi tantangan yang mendesak karena sifatnya yang dinamis dan sulit dikenali. Botnet modern mampu beroperasi dengan teknik komunikasi yang canggih seperti peer-to-peer (P2P), enkripsi lalu lintas, serta

penggunaan protokol yang menyerupai lalu lintas normal. Data dari laporan Cisco Annual Internet Report (2018 – 2023) menunjukkan bahwa lalu lintas jaringan global meningkat lebih dari 30% dalam tiga tahun terakhir, sehingga memperluas permukaan serangan dan meningkatkan kompleksitas dalam mengidentifikasi lalu lintas berbahaya (Cisco, 2020). Oleh karena itu, dibutuhkan pendekatan deteksi yang adaptif dan mampu mengidentifikasi pola anomali tersembunyi dalam data jaringan berskala besar (Abdurrasyid, Sitohang, et al., 2024).

Intrusion Detection System (IDS) menjadi salah satu solusi utama untuk mengidentifikasi aktivitas berbahaya dalam jaringan. IDS berbasis tanda tangan (*signature-based*) memang terbukti efektif dalam mendeteksi serangan yang sudah memiliki pola tertentu. Akan tetapi, pendekatan ini dianggap lemah dalam menghadapi serangan baru seperti *zero-day attack* atau variasi botnet yang belum pernah dikenali sebelumnya. Pendekatan konvensional yang umum digunakan selama ini mengandalkan *misuse detection* atau *signature matching*, yaitu membandingkan traffic jaringan dengan basis data signature serangan yang sudah diketahui, sehingga meski akurasi tinggi terhadap serangan dikenal, sistem ini sangat tergantung pada update *signature* dan gagal mendeteksi ancaman baru atau variasi yang berbeda dari pola yang sudah ada (misalnya serangan *zero-day* atau *polymorphic malware*) (Díaz-Verdejo et al., 2022; Maseno et al., 2022). Oleh karena itu, peneliti dan praktisi keamanan jaringan mulai beralih ke pendekatan berbasis anomali.

Pendekatan ini merupakan pendekatan yang tidak hanya mencakup penggunaan teknik statistik dan *rule-based* untuk mendefinisikan profil perilaku normal, tetapi juga berbagai metode pembelajaran mesin dan pembelajaran mendalam seperti clustering, autoencoder, satu-kelas SVM (*one class* SVM), LSTM atau model sekuensial untuk analisis waktu antar event, serta metode *semi-supervised* atau tidak berlabel untuk mendeteksi serangan yang belum pernah muncul sebelumnya (Moriano et al., 2025; Pinto et al., 2023). Karena keterbatasan metode *signature-based* dalam mendeteksi serangan baru dan varian serangan botnet yang belum dikenal, penulis memilih pendekatan pembelajaran mesin dan pembelajaran mendalam. Pendekatan ini memungkinkan model belajar dari data (termasuk pola perilaku dan fitur nondeterministik) sehingga lebih

adaptif terhadap ancaman yang berubah-ubah. Pembelajaran mendalam khususnya mampu mengekstrak representasi fitur secara otomatis dari data mentah, menangani data berdimensi tinggi dan non-linier dengan lebih baik, serta mengurangi kebutuhan *feature engineering* manual. Dengan pendekatan ini, sistem dapat lebih cepat mendeteksi *zero-day attack* atau serangan yang tidak memiliki signature sebelumnya, serta memiliki performa yang lebih baik dalam akurasi, presisi, dan deteksi anomali dibandingkan metode konvensional (Duan et al., 2022; Geetha & Brahmananda, 2025). Untuk mengevaluasi efektivitas model, dibutuhkan dataset yang representatif yang mampu mencerminkan variasi serangan modern secara nyata.

Salah satu dataset populer yang digunakan dalam penelitian keamanan jaringan adalah CIC-IDS-2018. Dataset ini dikembangkan oleh *Canadian Institute for Cybersecurity* dan berisi berbagai jenis lalu lintas normal serta berbahaya, termasuk botnet, DDoS, *brute-force*, dan serangan lainnya. Kelengkapan dan kompleksitas dataset ini menjadikannya sangat relevan dalam penelitian keamanan jaringan. Dengan menggunakan dataset ini, peneliti dapat menguji performa model deteksi secara komprehensif. Selain itu, dataset ini juga memungkinkan analisis lebih realistis karena menampilkan variasi trafik sesuai dengan kondisi dunia nyata. Hal ini menjadikannya standar acuan di bidang keamanan jaringan.

Beberapa pendekatan berbasis anomali yang telah digunakan saat ini untuk mendeteksi botnet adalah dengan menggunakan pembelajaran mesin, mulai dari algoritma klasifikasi tradisional seperti *Random Forest* dan *Support Vector Machine* hingga pendekatan berbasis *deep learning* (Hossain & Islam, 2023). Namun, metode yang telah disebutkan memiliki berbagai kelemahan seperti *Random Forest* yang sangat rentan terhadap waktu pelatihan dan inferensi tinggi jika memiliki jumlah pohon yang banyak, *Support Vector Machine* yang mengalami kesulitan skala pada dataset besar dengan fitur dimensi tinggi dan *deep learning* yang membutuhkan data pelatihan besar (Ali et al., 2025). Dalam konteks tersebut, pendekatan unsupervised learning menjadi relevan karena memungkinkan sistem mempelajari struktur dan pola utama dari data tanpa memerlukan label kelas. Pendekatan ini menitikberatkan pada pemodelan karakteristik lalu lintas normal, sehingga metode yang dapat merepresentasikan dan merekonstruksi data secara

efektif. Dari metode tersebut, autoencoder menjadi salah satu teknik yang menjanjikan. Autoencoder merupakan arsitektur jaringan saraf tiruan yang digunakan untuk merekonstruksi kembali input dengan mempelajari representasi tersembunyi (*latent representation*). Kemampuan ini membuat autoencoder efektif dalam mendeteksi anomali, termasuk pada trafik botnet, karena perbedaan signifikan antara pola normal dan pola anomali akan menghasilkan kesalahan rekonstruksi yang tinggi (De Bettignies, 2021; Dini et al., 2023).

Meski demikian, terdapat berbagai tantangan dalam implementasi autoencoder untuk deteksi botnet. Salah satunya adalah masalah ketidakseimbangan kelas pada dataset CIC IDS-2018, di mana jumlah data normal jauh lebih besar daripada trafik botnet. Ketidakseimbangan kelas ini dapat membuat model bias terhadap kelas mayoritas. Untuk itu penulis perlu menerapkan strategi pengimbangan data dan strategi training yang khusus agar model tetap sensitif terhadap kelas minoritas (serangan) tanpa menaikkan *false positive* secara berlebihan. Beberapa pendekatan yang dapat dilakukan untuk memastikan keseimbangan kelas adalah *oversampling*, *generative augmentation* dengan GAN, dan *one-class classification* (Shanmugam et al., 2024).

Penelitian ini menggunakan pendekatan *random oversampling*, yaitu teknik penyeimbangan data dengan menambah jumlah sampel pada kelas minoritas melalui duplikasi acak data serangan (Talukder et al., 2024). Pendekatan ini dipilih karena relatif sederhana, tidak memerlukan asumsi distribusi data tertentu, serta efektif dalam meningkatkan representasi kelas minoritas pada proses pelatihan model, sehingga autoencoder dapat mempelajari karakteristik lalu lintas botnet secara lebih proporsional (Saputra et al., 2024).

Selain itu, arsitektur autoencoder memerlukan optimasi parameter yang cermat agar dapat menangkap pola tersembunyi secara efektif. Tantangan tersebut muncul karena karakteristik data jaringan yang tidak seimbang, di mana jumlah lalu lintas normal jauh lebih banyak dibandingkan lalu lintas botnet. Ketidakseimbangan ini dapat menyebabkan model belajar lebih dominan pada pola normal sehingga kemampuan deteksinya terhadap lalu lintas anomali menjadi menurun (Ghosh et al., 2024). Selain itu, pemilihan arsitektur autoencoder, seperti jumlah lapisan tersembunyi, ukuran representasi laten, dan fungsi

aktivasi, sangat memengaruhi kualitas hasil rekonstruksi. Tidak adanya standar arsitektur yang optimal membuat setiap penelitian perlu melakukan eksplorasi konfigurasi yang berbeda-beda (Altalhan et al., 2025). Tantangan lain adalah pra pemrosesan data yang cukup kompleks. Pada beberapa kasus, performa model menurun ketika diuji pada variasi serangan yang lebih jarang muncul.

1.2 Rumusan Masalah

Berdasarkan uraian pada subbab sebelumnya, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan teknik *random oversampling* dalam menangani ketidakseimbangan kelas pada dataset CIC-IDS-2018 sebelum proses pelatihan model ?
2. Bagaimana pengaruh berbagai metode normalisasi data terhadap stabilitas dan konvergensi model *unsupervised learning* autoencoder ?
3. Bagaimana efektivitas model autoencoder yang diusulkan dalam mendeteksi serangan botnet berdasarkan metrik evaluasi ?

1.3 Tujuan

Tujuan dari penelitian ini adalah :

1. Mengevaluasi penerapan teknik *random oversampling* dalam mengatasi ketidakseimbangan kelas pada dataset CIC-IDS-2018.
2. Mengevaluasi pengaruh penerapan berbagai metode normalisasi data terhadap tingkat stabilitas dan konvergensi dalam pelatihan model *unsupervised learning* autoencoder.
3. Mengukur tingkat efektivitas model autoencoder yang diusulkan dalam mendeteksi serangan botnet menggunakan metrik evaluasi kuantitatif yang komprehensif.

1.4 Manfaat

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Memberikan kontribusi terhadap pengembangan keilmuan di bidang keamanan siber dan deteksi anomali, khususnya dalam penerapan metode *unsupervised*

learning autoencoder dan evaluasi teknik normalisasi data untuk mendeteksi lalu lintas botnet pada dataset berskala ekstrem.

2. Menjadi bahan rujukan bagi penelitian selanjutnya dalam pengembangan model *Intrusion Detection System* (IDS) yang stabil, adaptif, dan tangguh dalam menangani masalah ketidakseimbangan kelas pada lalu lintas jaringan modern.
3. Memberikan pemahaman serta dukungan konseptual sebagai solusi jangka pendek bagi praktisi keamanan siber dan administrator jaringan dalam menentukan konfigurasi pra-pemrosesan data yang optimal, sekaligus membantu merancang sistem penyaring lalu lintas (*traffic filter*) awal yang efektif dalam meminimalkan alarm palsu (*false positive*) saat menghadapi serangan jaringan.

1.5 Ruang Lingkup

Ruang lingkup penelitian ini ditetapkan untuk memperjelas batasan, fokus, dan tahapan penelitian agar pelaksanaannya tetap terarah serta selaras dengan tujuan yang telah ditetapkan. Adapun ruang lingkup penelitian ini meliputi hal-hal sebagai berikut:

1. Penelitian ini dibatasi pada deteksi anomali lalu lintas jaringan, sehingga analisis hanya difokuskan pada identifikasi dan pemisahan antara lalu lintas jaringan normal (*benign*) dan lalu lintas serangan botnet.
2. Data yang digunakan diperoleh dari dataset sekunder publik CIC-IDS-2018 edisi 02 Maret 2018, dengan variabel yang dianalisis meliputi 41 fitur jaringan hasil seleksi seperti *Flow Duration*, *Tot Fwd Pkts*, dan *Pkt Len Mean* beserta *Label* dari setiap baris data.
3. Metode yang diterapkan dalam penelitian ini adalah *unsupervised learning* autoencoder untuk deteksi anomali, yang dipadukan dengan teknik *Random Oversampling* untuk penanganan ketidakseimbangan kelas serta pencegahan dominasi nilai ekstrem dengan tiga metode normalisasi (Min-Max, Standardization, dan Robust).
4. Pelaksanaan penelitian meliputi beberapa tahapan, dimulai dari persiapan data, yaitu transformasi label, seleksi fitur, penyeimbangan kelas, pembersihan nilai tidak valid (*infinite*), normalisasi, serta pembagian data latih dan data uji dengan rasio 70:30. Tahap pemodelan mencakup pelatihan arsitektur autoencoder secara

eksklusif pada data normal, perhitungan *Mean Squared Error* (MSE), penentuan ambang batas (*threshold*) menggunakan persentil ke-95, serta pengujian hasil model menggunakan *confusion matrix* dan kurva ROC-AUC untuk memberikan evaluasi performa deteksi yang terukur.

1.6 Kebaruan

Kebaruan dalam penelitian ini terletak pada penerapan metode unsupervised learning berbasis autoencoder untuk mendeteksi lalu lintas botnet pada dataset CIC-IDS-2018 yang merepresentasikan kondisi jaringan modern dan kompleks. Berbeda dengan penelitian sebelumnya yang umumnya menggunakan pendekatan *supervised learning* seperti *Support Vector Machine* (SVM) atau *Random Forest* yang membutuhkan data berlabel, penelitian ini mengandalkan pembelajaran tanpa label agar model mampu mengenali pola anomali secara otomatis melalui proses rekonstruksi fitur jaringan. Nilai *reconstruction error* yang dihasilkan digunakan untuk menentukan ambang batas (*threshold*) deteksi anomali secara adaptif sesuai distribusi data.

1.7 Sistematika Penulisan

Terdapat sistematika penelitian yang ditulis sebagai alur dari penelitian, yaitu :

1. BAB I Pendahuluan

   Penjelasan mengenai latar belakang, rumusan masalah, terdapat pula tujuan, disertai dengan manfaat penelitian, ruang lingkup masalah, dan kebaruan pada penelitian.
2. BAB II Tinjauan Pustaka

   Penjelasan terkait bagian dari bab ini mengenai teori-teori dasar pada penelitian yang ada sebelumnya, menyertakan berbagai referensi terkait, dan penemuan penelitian yang mirip dan juga relevan.
3. BAB III Metode Penelitian

   Terdapat paparan informasi mengenai waktu, tahapan penelitian, teknik yang digunakan atau yang berkaitan terhadap penelitian yang dijalankan, dan metode terpilih serta dipakai dalam proses penelitian.
4. BAB IV Hasil dan Pembahasan

Bab ini berisikan penjelasan terkait rekapan hasil penelitian yang telah dilakukan, yang berisi penjelasan bahasan mengenai implementasi proses yang telah dijalankan.

5. BAB V Penutup

   Bagian ini terdapat Penutup yang berisi bahasan mengenai kesimpulan-kesimpulan pada penelitian yang telah berhasil dilakukan, beserta dengan kesimpulan dari perbandingan penggunaan model pada penelitian, dan disertai dengan saran.