BAB I
PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam dokumen skripsi ini yang berjudul Optimasi *Intrusion Detection System* melalui Seleksi Fitur Berbasis *Pearson Correlation* dan Model BiGRU dengan *Rule-Based Filtering.*

1.1. Latar Belakang

Perkembangan teknologi internet yang pesat, khususnya dengan kemunculan *Internet of Things* (IoT), telah menciptakan eksosistem global yang menghubungkan miliaran perangkat fisik ke jaringan digital. Kemunculan IoT memungkinkan tingkat otomatisasi dan interkonektivitas yang belum pernah terjadi sebelumnya, di mana perangkat dapat mengumpulkan, menukar, dan memproses data secara *real-time* tanpa intervensi manusia. Transformasi ini memungkinkan tingkat otomatisasi yang belum pernah terjadi sebelumnya di berbagai sektor strategis, seperti layanan kesehatan cerdas (*smart healthcare*), rumah cerdas (*smart homes*), transportasi cerdas (*smart transportation*), dan pendidikan cerdas (*smart education*) (Al-Garadi dkk., 2020).

Namun, meskipun membawa manfaat yang besar, transformasi digital yang ada juga menimbulkan tantangan baru dalam hal keamanan dan privasi. Banyak perangkat IoT beroperasi dengan sumber daya terbatas dan menggunakan protokol keamanan yang minimal, sehingga menjadi rentan terhadap serangan siber seperti *Distributed Denial of Service* (DDoS). Dalam serangan ini, ribuan hingga jutaan perangkat IoT yang telah dikompromikan dapat mengirimkan sejumlah besar permintaan ke suatu target dalam waktu singkat, sehingga menyebabkan kelebihan kapasitas server dan membuat layanan menjadi tidak tersedia. Ketidaktersediaan ini akan mencegah lalu lintas yang sah antara server dan klien (Rohit dkk., 2019).

Sebagai mekanisme pertahanan pertama terhadap ancaman tersebut, *Signature-based Intrusion Detection System* (SIDS) telah lama diadopsi sebagai metode konvensional dalam keamanan jaringan. SIDS bekerja menggunakan mekanisme pencocokan pola (*pattern matching*), yang membandingkan lalu lintas jaringan yang masuk dengan basis data

signature serangan yang telah diketahui. Metode ini dinilai sangat efektif dan memiliki akurasi yang tinggi untuk intrusi yang telah diketahui sebelumnya (Khraisat dkk., 2019). Namun, ketergantungan SIDS pada pembaruan basis data secara manual menjadi kelemahan fundamental, terutama dalam ekosistem IoT yang dinamis. SIDS tidak mampu mendeteksi serangan varian baru, serangan polimorfik, atau serangan *zero-day* (serangan yang belum pernah diketahui sebelumnya) karena pola serangan tersebut belum tersedia dalam basis data sistem. Keterbatasan ini membuat metode berbasis *signature* menjadi kurang efektif menghadapi ancaman modern yang berevolusi dengan cepat (Buczak & Guven, 2016).

Keterbatasan SIDS dalam menangani serangan yang berevolusi, mendorong pergeseran pendekatan menjadi *Anomaly-based Intrusion Detection System* (AIDS). AIDS mampu untuk mengidentifikasi serangan *zero-day* karena metode ini dapat mengenali aktivitas pengguna yang abnormal tanpa bergantung pada basis data *signature*. AIDS dapat memicu peringatan atau sinyal bahaya ketika perilaku yang diperiksa berbeda dari perilaku biasa (Khraisat dkk., 2019). Karakteristik tersebut menjadikan AIDS sebagai pendekatan yang jauh lebih relevan dan adaptif untuk lingkungan IoT, mengingat jenis serangan terus berkembang pesat dan sering kali belum tedokumentasikan.

Penelitian lanjutan dalam pengembangan AIDS menjadi sangat esensial, khususnya melalui penerapan teknik ML dan DL. Beberapa penelitian terdahulu telah menerapkan algoritma ML konvensional seperti *K-Nearest Neighbors* (KNN), *Support Vector Machines* (SVM), dan *Random Forest* untuk mendeteksi intrusi pada jaringan IoT. Meskipun algoritma ini menunjukkan hasil yang cukup baik pada dataset standar, performanya cenderung menurun ketika dihadapkan pada volume data yang masif dan berdimensi tinggi seperti pada dataset Bot-IoT (Al-Garadi dkk., 2020).

Sebagai solusi atas keterbatasan tersebut, teknik DL mulai banyak diadopsi karena kemampuannya dalam mengekstraksi fitur secara otomatis (Sharma & Jain, 2019) dan menangani data non-linear yang kompleks pada data berskala besar (Maggu dkk., 2025). Model seperti *Multilayer Perceptron* (MLP), *Convolutional Neural Networks* (CNN), dan *Long Short-Term Memory* (LSTM) terbukti mampu mengungguli metode ML tradisional dalam hal akurasi deteksi dan kemampuan generalisasi terhadap serangan baru (Chakrabarti & Saha, 2019 ; Koroniotis dkk., 2019). Dalam penelitian yang dilakukan oleh Koroniotis dkk. (2019), model berbasis LSTM menunjukkan performa lebih baik dibanding SVM pada

dataset Bot-IoT, sehingga menegaskan efektivitas DL dalam mengidentifikasikan *traffic botnet.* Namun, penerapan DL pada dataset IoT berskala besar seperti Bot-IoT juga menimbulkan tantangan baru, antara lain beban komputasi yang tinggi dan risiko *overfitting* akibat dimensi fitur yang besar.

Untuk mengatasi tantangan dimensi data yang tinggi dan beban komputasi yang besar, penerapan teknik seleksi fitur (*feature selection*) sebelum melakukan proses pelatihan model menjadi langkah awal yang krusial. Penelitian oleh Shafiq dkk. (2020) pada dataset Bot-IoT menunjukkan bahwa seleksi fitur yang efektif mampu mereduksi kompleksitas data secara signifikan tanpa mengorbankan akurasi deteksi, sekaligus mempercepat waktu pelatihan model. Selain itu, proses seleksi fitur juga penting untuk mengidentifikasi atribut-atribut utama yang berkontribusi terhadap deteksi intrusi. Dalam konteks ini, interpretabilitas menjadi sangat penting karena memungkinkan analis keamanan untuk memahami mengapa fitur-fitur tertentu dianggap relevan untuk deteksi intrusi. Dengan mengetahui fitur-fitur yang paling berkontribusi terhadap proses deteksi, pakar keamanan dapat memvalidasi relevansinya berdasarkan pengetahuan domain, membangun kepercayaan terhadap sistem, serta menjadikannya sebagai referensi pendukung dalam perumusan kebijakan keamanan dan pengembangan aturan mitigasi yang lebih spesifik (Shyaa dkk., 2024). Di samping seleksi fitur, optimasi *hyperparameter* pada model DL juga sangat penting karena konfigurasi yang tidak tepat dapat menyebabkan model terjebak pada *local optima.*

Penelitian terdahulu telah banyak mengeksplorasi penerapan algoritma DL untuk klasifikasi biner *botnet*. Kamal & Mashaly (2025), melakukan penelitian menggunakan kombinasi arsitektur CNN dan MLP untuk klasifikasi biner pada dataset IoT-23 dan NF-Bot-IoT-v2. Dari hasil penelitian tersebut, didapat akurasi mencapai 99,94% pada dataset IoT-23 dan 99,96% pada dataset NF-Bot-IoT-v2. Di sisi lain, Ullah dkk. (2021) juga melakukan penelitian menggunakan kombinasi arsitektur CNN dan GRU untuk klasifikasi biner pada dataset IoT-DS-2 dengan hasil akurasi, *precision*, *recall*, dan *f1-score* yang tinggi, yaitu lebih dari 99,50%, yang menandakan prakiraan FP (*False Positive*) dan FN (*False Negative*) yang rendah. Selain pada dataset-dataset tersebut, eksplorasi mendalam juga telah dilakukan secara spesifik pada dataset lain. Penelitian oleh (Roopak dkk., 2019) mengusulkan dan membandingkan empat arsitektur DL, yaitu MLP, CNN, LSTM, dan hybrid CNN+LSTM, untuk mendeteksi serangan DDoS pada jaringan IoT menggunakan dataset CICIDS2017. Hasil penelitian menunjukkan bahwa model CNN+LSTM

memberikan performa terbaik dengan akurasi 97,16%, mengungguli model DL lain serta algoritma ML tradisional seperti SVM dan Random Forest.

Meskipun hasil pada dataset-dataset tersebut sangat memuaskan, dataset Bot-IoT kini menjadi *benchmark* utama dan sangat relevan dalam penelitian IDS karena skalanya yang besar dan kompleksitas serangan yang tinggi. Ullah & Mahmoud (2022) melakukan evaluasi komprehensif menggunakan berbagai algoritma DL untuk deteksi kelas biner pada dataset ini. Temuan mereka menunjukkan bahwa dataset Bot-IoT sangat ideal untuk menguji performa model DL, di mana model BiLSTM mencapai akurasi tertinggi sebesar 99,96%, diikuti oleh GRU (99,93%) dan LSTM (99,90%). Koroniotis dkk. (2019), pengembang dataset Bot-IoT, memvalidasi kualitas dataset dengan mengevaluasi performa model DL berbasis urutan waktu, yakni LSTM dan RNN, menggunakan 10 fitur terbaik yang dipilih berdasarkan *correlation coefficient* dan *information gain*. Hasil evaluasi tersebut menunjukkan kinerja yang sangat baik, dengan akurasi masing-masing sebesar 99,74% untuk LSTM dan 99,55% untuk RNN, yang menegaskan bahwa pendeketan berbasis urutan waktu sangat efektif dalam menedeteksi aktivitas *botnet*.

Meskipun demikian, masalah mendasar pada dataset Bot-IoT adalah dimensi data yang tinggi dan ketidakseimbangan kelas yang ekstrem. Pendekatan konvensional yang mengandalkan *oversampling* seringkali menambah beban komputasi dan risiko *noise*. Oleh karena itu, penelitian terkini cenderung mengusulkan strategi efisiensi melalui seleksi fitur yang ketat. Adi dkk (2025) mengusulkan metode seleksi fitur berbasis interkorelasi antar-fitur dengan batasan *threshold* dinamis untuk menentukan jumlah fitur optimal guna mengatasi performa rendah pada kelas minoritas. Sejalan dengan perkembangan tersebut, Xiang dkk. (2024) mengusulkan arsitektur ResNest-BiGRU untuk sistem deteksi intrusi pada lingkungan IoT, yang memanfaatkan kemampuan Bidirectional GRU (BiGRU) dalam memproses informasi sekuensial dari dua arah secara simultan. Pendekatan *bidirectional* ini memungkinkan model menangkap pola temporal secara lebih komprehensif, dengan akurasi mencapai 99,90% dan *f1-score* mencapai 99,88% pada dataset IoT.

Melihat kemajuan ini, penelitian ini mengusulkan integrasi seleksi fitur menggunakan gabungan metode *Pearson Correlation Coefficient* (PCC) untuk mengukur hubungan linear terkuat antara fitur dan target (label), serta perhitungan standar deviasi, dengan arsitektur BiGRU untuk deteksi intrusi biner pada dataset Bot-IoT. Pendekatan ini tidak hanya

bertujuan untuk meningkatkan performa prediksi, tetapi juga meningkatkan interpretabilitas model dengan memfokuskan pada atribut-atribut paling signifikan, sehingga diharapkan mamapu mengoptimalkan kemampuan deteksi intrusi. Dengan membatasi model pada fitur yang paling representatif, penelitian ini memungkinkan analisis transparan terhadap faktor lalu lintas jaringan yang memengaruhi deteksi, mengoptimalkan efisiensi model melalui reduksi dimensi data yang terkontrol, menekan kompleksitas komputasi, menjaga stabilitas pembelajaran, dan meningkatkan kemampuan generalisasi tanpa mengorbankan akurasi.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah bagaimana efektivitas dan performa model DL berbasis arsitektur BiGRU dalam melakukan klasifikasi biner pada dataset Bot-IoT dengan penerapan seleksi fitur menggunakan gabungan metode PCC untuk mengukur hubungan linear terkuat antara fitur dan target (label) serta perhitungan standar deviasi, serta apakah pendekatan tersebut mampu menghasilkan fitur yang paling relevan dan interpretatif sehingga dapat meningkatkan kualitas deteksi sekaligus mendukung pemebentukan mekanisme berbasis aturan.

1.3. Tujuan dan Manfaat

Tujuan dari penelitian ini adalah untuk mengembangkan dan mengevaluasi model klasifikasi biner berbasis arsitektur BiGRU dalam mendeteksi intrusi pada dataset Bot-IoT dengan menerapkan pendekatan seleksi fitur yang mengombinasikan PCC dan analisis standar deviasi untuk menentukan jumlah fitur optimal. Penelitian ini juga bertujuan untuk mengkaji apakah fitur terpilih tidak hanya mampu mempertahankan performa klasifikasi, tetapi juga memiliki tingkat interpretabilitas yang memadai sehingga dapat dimanfaatkan dalam perancangan mekanisme *filtering* berbasis aturan sebagai bagian dari pendekatan *hybrid* IDS.

Manfaat yang diharapkan dari penelitian ini adalah tersedianya model klasifikasi biner yang teruji secara eksperimental dalam mendeteksi intrusi pada lingkungan IoT dengan efisiensi komputasi yang lebih baik melalui reduksi dimensi fitur. Selain itu, penelitian ini memberikan kontribusi metodologis berupa pendekatan seleksi fitur berbasis kombinasi analisis korelasi dan perubahan standar deviasi untuk menentukan fitur yang paling relevan dan informatif. Fitur terpilih yang bersifat interpretatif juga berpotensi dimanfaatkan sebagai dasar pembentukan aturan dalam mekanisme *filtering* awal, sehingga mendukung

pengembangan sistem IDS yang lebih adaptif, efisien, dan aplikatif pada lingkungan IoT dengan keterbatasan sumber daya komputasi.

1.4. Ruang Lingkup

Ruang lingkup penelitian ini meliputi pengembangan dan evaluasi model deteksi anomali jaringan berbasis DL dengan fokus pada dataset Bot-IoT. Proses penelitian dibatasi pada tahap prapemrosesan data, seleksi fitur menggunakan penggabungan metode PCC dan perhitungan standar deviasi untuk menentukan jumlah fitur optimal, pelatihan model DL berbasis arsitektur BiGRU, serta evaluasi performa klasifikasi biner menggunakan metrik *precision*, *recall*, *F1-score*, dan analisis *confusion matrix*, termasuk perbandingan model dengan penelitian terdahulu.

Selain itu, penelitian ini mencakup perancangan mekanisme filtering awal berbasis aturan yang disusun berdasarkan interpretabilitas fitur hasil seleksi. Aturan tersebut diterapkan pada level eksperimen sebagai lapisan prapemrosesan sebelum inferensi model, guna mengevaluasi pendekatan *hybrid* antara *rule-based filtering* dan DL. Penelitian ini tidak mencakup implementasi sistem secara langsung pada lingkungan IDS *real-time*, integrasi dengan perangkat *firewall* nyata, maupun optimasi terhadap arsitektur jaringan komputer atau infrastruktur jaringan secara langsung.

1.5. Sistematika Penulisan

Struktur penulisan dalam laporan ini dibagi menjadi lima bab utama yang membahas penelitian "Optimasi *Intrusion Detection System* melalui Seleksi Fitur Berbasis *Pearson Correlation* dan Model BiGRU dengan *Rule-Based Filtering*". Pembagian ini disusun dengan tujuan agar penulisan menjadi lebih sistematis dan mudah dipahami. Berikut ini adalah ringkasan singkat dari setiap bab yang dibahas:

BAB I PENDAHULUAN

Bab ini menyajikan latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA

Bab ini menguraikan landasan teori yang mendukung penelitian serta kajian terhadap penelitian-penelitian terdahulu yang relevan dengan topik yang dibahas.

BAB III METODE PENELITIAN

Bab ini menjelaskan metode dan tahapan-tahapan penelitian yang dilakukan secara sistematis, mulai dari pengumpulan data, pengolahan data, perancangan model, hingga evaluasi hasil.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari implementasi metode yang diusulkan serta pembahasan dan analisis terhadap hasil yang diperoleh.

BAB V KESIMPULAN

Bab ini berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.