BAB I

PENDAHULUAN

1.1 Latar Belakang

Internet of Things (IoT) telah mengubah cara manusia berinteraksi dengan teknologi, menghubungkan berbagai perangkat fisik seperti kamera keamanan, *smart home devices*, dan perangkat lainnya dalam satu lingkungan digital yang terintegrasi. Perkembangan ini tidak hanya meningkatkan efisiensi aktivitas sehari-hari, tetapi juga memberikan peluang besar untuk meningkatkan produktivitas di berbagai sektor penting seperti industri, transportasi, dan kesehatan (Quy dkk., 2022). Seiring dengan peningkatan adopsi IoT, jumlah perangkat yang terhubung ke jaringan juga mengalami pertumbuhan yang pesat. Laporan pasar global menunjukkan bahwa jumlah perangkat IoT mencapai 18,5 miliar pada tahun 2024 dan diproyeksikan meningkat menjadi 21,1 miliar pada akhir tahun 2025 (IoT Analytics, 2025).

Peningkatan jumlah perangkat IoT berdampak langsung terhadap meningkatnya volume trafik jaringan. Kondisi ini turut meningkatkan risiko keamanan siber, karena perangkat IoT umumnya memiliki keterbatasan sumber daya komputasi, seperti kapasitas CPU dan memori yang terbatas, sehingga tidak mampu menerapkan mekanisme keamanan konvensional yang kompleks (Al-Garadi dkk., 2020). Kerentanan tersebut menjadikan jaringan IoT sebagai target serangan malware yang dapat menyebabkan gangguan layanan, pencurian data, hingga penyalahgunaan perangkat dalam skala besar. Oleh karena itu, dibutuhkan

sistem deteksi malware yang tidak hanya akurat, tetapi juga memiliki tingkat kesalahan yang rendah, sebagaimana ditunjukkan oleh penelitian (Hartinah dkk., 2023) yang memperoleh akurasi rata-rata 94% pada data latih dengan error rate tertinggi sebesar 10%, sehingga menunjukkan model mampu bekerja secara stabil dalam proses deteksi.

Pendekatan yang umum digunakan dalam deteksi malware IoT adalah analisis data trafik jaringan. Namun, data trafik IoT memiliki karakteristik berdimensi tinggi *(high dimensionality),* yaitu kondisi ketika dataset terdiri dari puluhan hingga ratusan fitur numerik hasil ekstraksi statistik jaringan Jumlah fitur yang besar tersebut memperluas ruang vektor fitur dan meningkatkan kompleksitas komputasi dalam proses pelatihan model. Survei yang dilakukan (De Keersmaeker dkk., 2023) menunjukkan bahwa banyak dataset IoT publik mengandung fitur statistik yang beragam dan dalam jumlah besar, sehingga memerlukan tahap prapemrosesan sebelum digunakan dalam pemodelan. (Sarwar dkk., 2025) menyatakan bahwa *high dimensionality* merupakan permasalahan umum dalam sistem deteksi anomali karena banyaknya fitur dapat menurunkan efisiensi komputasi dan memperberat proses deteksi apabila tidak disertai dengan teknik seleksi fitur atau reduksi dimensi. Selain itu, keberadaan fitur yang tidak relevan atau redundan juga berpotensi menurunkan kinerja model dan meningkatkan risiko overfitting (Westphal dkk., 2025).

Untuk mengatasi permasalahan tersebut, penelitian sebelumnya telah menerapkan teknik reduksi dimensi seperti *Mutual Information* (MI) dan *Principal Component Analysis* (PCA) untuk meningkatkan efisiensi dan kinerja deteksi

malware pada jaringan IoT (Al-Sarem dkk., 2022). Hasil penelitian tersebut menunjukkan bahwa MI dan PCA mampu menurunkan beban komputasi dan meningkatkan akurasi deteksi. Namun demikian, penelitian yang mengombinasikan MI dan PCA umumnya difokuskan pada klasifikasi konvensional seperti Decision Tree, SVM, atau Random Forest serta diuji pada satu dataset tertentu (Kaushik dkk., 2022). Pendekatan tersebut belum banyak dianalisis pada arsitektur berbasis sekuensial yang memodelkan pola temporal trafik jaringan, khususnya dalam skenario evaluasi multi-dataset IoT yang dilengkapi dengan validasi statistik. Oleh karena itu, diperlukan evaluasi komparatif yang sistematis untuk memahami efektivitas reduksi dimensi terhadap model sekuensial dalam konteks deteksi malware IoT.

Sebagian penelitian sebelumnya belum secara eksplisit memvalidasi perbedaan performa model setelah penerapan reduksi dimensi menggunakan pengujian statistik. Padahal, perbedaan nilai performa yang ditunjukan oleh metrik evaluasi tidak selalu mencerminkan peningkatan yang signifikan secara statistik. Oleh karena itu, diperlukan pendekatan evaluasi yang mampu memastikan bahwa perubahan performa yang dihasilkan benar-benar bermakna dan bukan sekedar variasi acak.

Di sisi lain, deteksi malware pada jaringan IoT tidak hanya bergantung pada karakteristik statis data, tetapi juga pada pola temporal yang muncul dalam trafik jaringan. Serangan malware umumnya terjadi dalam bentuk urutan kejadian tertentu, sehingga memerlukan model yang mampu memodelkan data sekuensial. Model berbasis urutan seperti Gated Recurrent Unit (GRU) memiliki kemampuan

untuk menangkap pola temporal dengan kompleksitas komputasi yang relatif lebih rendah dibandingkan arsitektur sekuensial lain seperti LSTM. Karakteristik ini menjadikan GRU lebih sesuai untuk lingkungan IoT yang memiliki keterbatasan sumber daya komputasi. Meskipun demikian, pengaruh penerapan teknik reduksi dimensi terhadap performa arsitektur GRU pada data trafik IoT berdimensi tinggi masih memerlukan analisis yang sistematis.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pendekatan deteksi malware pada jaringan IoT dengan mengombinasikan teknik reduksi dimensi hybrid (MI dan PCA) dengan GRU. Pendekatan hybrid digunakan untuk memanfaatkan keunggulan MI dalam memilih fitur yang relevan serta keunggulan PCA dalam mereduksi dimensi data dengan tetap mempertahankan informasi utama. Selanjutnya pemilihan GRU bertujuan untuk menangkap pola temporal pada trafik jaringan tanpa menambah beban komputasi yang signifikan. Untuk mengevaluasi kemampuan generalisasi metode yang diusulkan, penelitian ini diuji menggunakan tiga dataset dengan karakteristik yang berbeda, yaitu N-BaIoT, CIC-IoT2023, dan Edge-IIoTset. N-BaIoT dipilih karena memiliki dimensi fitur yang tinggi dan merepresentasikan kondisi *high-dimensional* data, CIC-IoT23 digunakan karena merupakan dataset IoT terbaru yang mencakup berbagai jenis serangan dan mencerminkan kondisi ancaman terkini. Sedangkan Edge IIoTset mencerminkan skenario IoT dan Industrial IoT (IIoT) yang lebih realistis (Ferrag dkk., 2022). Penggunaan tiga dataset dengan karakteristik berbeda ini bertujuan untuk menguji konsistensi serta kemampuan generalisasi pendekatan yang diusulkan pada berbagai skenario trafik IoT.

Untuk memastikan bahwa perbedaan performa yang dihasilkan bersifat signifikan secara statistik, McNemar dipilih karena merupakan uji non-parametrik yang dirancang untuk data kategorikal berpasangan, dimana dua model diuji pada dataset yang sama. Dalam konteks penelitian ini, uji McNemar digunakan untuk menganalisis perbedaan keputusan klasifikasi antara model baseline dan model dengan penerapan reduksi dimensi. Fokus uji ini terletak pada kasus ketidaksepakatan *(disagreement)* antar model, sehingga mampu memberikan evaluasi yang lebih tepat terhadap perubahan performa akibat penerapan metode yang diusulkan dibandingkan dengan uji statistik lain yang berfokus pada rata-rata nilai performa. Dengan demikian, uji McNemar memungkinkan evaluasi perbedaan performa pada tingkat prediksi individual, sehingga memberikan validasi yang lebih kuat terhadap efektivitas metode yang diusulkan.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, diperoleh permasalahan pada penelitian ini, diantaranya :

a. Bagaimana pengaruh penerapan teknik reduksi dimensi PCA, MI dan Hybrid (MI - PCA) terhadap performa model GRU dalam mendeteksi malware pada jaringan IoT?
b. Apakah perbedaan performa model deteksi malware dengan dan tanpa teknik reduksi dimensi bersifat signifikan secara statistik berdasarkan uji McNemar?

1.3 Tujuan Penelitian

Tujuan penelitian ini dapat dirumuskan sebagai berikut :

a. Menganalisis pengaruh penerapan teknik reduksi dimensi PCA, MI dan Hybrid (MI-PCA) terhadap performa model GRU dalam mendeteksi malware pada jaringan IoT.
b. Menguji signifikansi perbedaan performa deteksi malware antara model GRU tanpa reduksi dimensi dan model GRU dengan penerapan teknik resuksi dimensi berdasarkan uji McNemar.

1.4 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini, sebagai berikut:

a. Memberikan kontribusi pada pengembangan ilmu pengetahuan di bidang keamanan siber khususnya dalam deteksi malware.
b. Menyediakan referensi baru mengenai pengaruh teknik reduksi dimensi (PCA, MI, dan Hybrid) terhadap kinerja model GRU dalam deteksi malware pada dataset IoT.
c. Menyediakan analisis validasi statistik mengenai efektivitas pengurangan fitur menggunakan McNemar untuk mengukur signifikansi perbedaan performa dalam model deep learning, khususnya dalam konteks deteksi malware IoT.

1.5 Batasan Masalah

a. Penelitian ini difokuskan pada deteksi serangan malware pada jaringan IoT menggunakan data trafik jaringan, tanpa membahas jenis serangan siber lainnya.
b. Penelitian ini hanya mengevaluasi model GRU *(Gated Recurrent Unit)* sebagai arsitektur utama, dengan 3 reduksi dimensi yaitu PCA, MI dan Hybrid. Penelitian tidak mengeksplorasi arsitektur lain seperti CNN, LSTM, atau Transformer
c. Dataset yang digunakan adalah N-BaloT, CIC-IoT2023 dan Edge-IIoTset. Dataset ini dipilih karena merepresentasikan berbagai karakteristik trafik IoT dan jenis serangan yang berbeda.
d. Evaluasi model dibatasi pada lingkungan offline menggunakan perangkat laptop pribadi, dan hanya mencakup metrik kinerja seperti akurasi, presisi, recall, F1-score serta specificity. Model tidak diuji dalam lingkungan *real-time* atau perangkat IoT secara langsung.
e. Analisis validasi statistik dibatasi pada penerapan McNemar, yang digunakan untuk mengukur signifikansi perbedaan performa antara model baseline dan model dengan penerapan reduksi dimensi (PCA, MI, dan Hybrid).

1.6 Sistematika Penulisan

Sistematika penulisan digunakan supaya penulisan laporan penelitian dapat terarah dan tersusun sesuai tahapan penelitian. Sistematika penulisan dalam penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang penelitian, rumusan masalah yang dikaji, tujuan penelitian, manfaat penelitian, metodologi yang digunakan, serta sistematika penulisan dalam pelaporan penelitian.

BAB II LANDASAN TEORI

Bab ini berisi pembahasan teori-teori yang berkaitan dengan penelitian, termasuk matriks penelitian, konsep, teknik, dan algoritma yang digunakan. Selain itu, bab ini juga memaparkan tinjauan penelitian sebelumnya yang serupa serta menjelaskan aspek keterbaruan dari penelitian yang dilakukan.

BAB III METODOLOGI PENELITIAN

Bab ini berisi uraian metode penelitian yang digunakan, termasuk posisi topik penelitian dalam *roadmap* penelitian, objek penelitian, variabel penelitian, serta tahapan-tahapan yang dilakukan dalam penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil penelitian serta pembahasan terkait perancangan yang telah dijelaskan pada bab sebelumnya. Pembahasan mencakup pembuatan rancangan model dan algoritma yang digunakan, serta eksperimen yang dilakukan

bersama model lainnya untuk mengevaluasi kinerja dan efektivitas pendekatan yang diterapkan.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil eksperimen yang telah dilakukan dan saran untuk penelitian selanjutnya berdasarkan batasan dan temuan dalam penelitian yang membahas topik serupa.