

BAB I

PENDAHULUAN

A. Latar Belakang

Perkembangan teknologi informasi yang pesat dalam beberapa dekade terakhir telah mendorong pemanfaatan jaringan komputer secara luas di berbagai bidang, mulai dari pemerintahan, industri, hingga pendidikan. Jaringan komputer memungkinkan pertukaran data dan komunikasi yang efisien, namun juga membuka celah bagi berbagai ancaman keamanan [1]. Intrusi jaringan merupakan upaya pihak tidak bertanggung jawab untuk mengakses, merusak, atau mencuri data dalam sistem jaringan secara ilegal.

Sistem keamanan jaringan sendiri adalah sekumpulan kebijakan dan mekanisme teknis yang dirancang untuk melindungi integritas, kerahasiaan, dan ketersediaan data serta layanan yang berjalan di dalamnya. Dalam era digitalisasi yang terus berkembang pesat, hampir seluruh sektor di Indonesia telah mengadopsi sistem berbasis jaringan untuk mendukung kegiatan operasional, seperti layanan perbankan digital, *e-government*, sistem pembelajaran daring, dan layanan publik berbasis aplikasi [1]. Namun, seiring dengan peningkatan adopsi teknologi informasi, potensi terjadinya serangan siber juga semakin tinggi dan kompleks. Serangan siber dapat menyebabkan berbagai dampak negatif, mulai dari kerugian finansial, pencurian data sensitif, hingga gangguan operasional layanan penting.

Berdasarkan laporan dari Badan Siber dan Sandi Negara, Indonesia mencatat ratusan kasus serangan siber selama tahun 2023, yang meliputi serangan seperti *Ransomware*, *Phishing*, *Distributed Denial of Service (DoS)*, dan *Advanced Persistent Threat (APT)* [2], [3]. Serangan-serangan ini tidak hanya menasar perusahaan besar, tetapi juga menargetkan instansi pemerintah, sektor kesehatan, pendidikan, dan keuangan. Meningkatnya kompleksitas dan frekuensi serangan siber menuntut peningkatan sistem pertahanan jaringan yang tidak hanya bersifat reaktif, tetapi juga proaktif dan adaptif.

Salah satu sistem yang umum digunakan dalam pengamanan jaringan adalah Intrusion Detection System (IDS) yang berfungsi untuk memantau lalu lintas jaringan dan mengidentifikasi aktivitas mencurigakan yang dapat mengindikasikan adanya serangan [4]. Meskipun sistem IDS konvensional berbasis tanda tangan atau pola serangan diketahui cukup efektif untuk mengidentifikasi ancaman yang telah dikenal, sistem ini masih memiliki keterbatasan dalam mengidentifikasi serangan baru (*zero-day attack*). Lebih krusial lagi, sistem IDS tradisional seringkali hanya memberikan indikasi adanya intrusi tanpa mampu mengklasifikasikan jenis serangan secara spesifik, sehingga mempersulit respons cepat, terarah, dan tepat dari administrator jaringan. Kemampuan untuk mengetahui secara pasti jenis intrusi yang terjadi (misalnya, apakah itu *DoS*, *malware*, atau *probe*) sangat vital untuk mitigasi yang efektif.

Untuk mengatasi keterbatasan dalam kemampuan klasifikasi intrusi jaringan tersebut, pendekatan berbasis deep learning mulai dikembangkan dalam sistem keamanan jaringan. *Deep learning* merupakan cabang dari *machine learning* yang memiliki kemampuan untuk mengenali pola kompleks dari data dalam jumlah besar secara otomatis melalui jaringan saraf berlapis-lapis (*deep neural network*) [5]. Pendekatan ini telah terbukti unggul dalam berbagai aplikasi yang memerlukan pengenalan pola yang canggih, seperti pengolahan citra, suara, dan teks. Dalam konteks keamanan jaringan, algoritma *deep learning* menawarkan keunggulan dalam menganalisis lalu lintas jaringan dan mengenali anomali, serta mampu mengklasifikasikan berbagai jenis intrusi dengan presisi yang lebih tinggi dibandingkan dengan metode konvensional [6]. Keunggulan utama dari pendekatan ini adalah kemampuannya untuk belajar dari data tidak terstruktur dan melakukan klasifikasi terhadap pola yang terus berubah secara dinamis, bahkan untuk serangan yang belum pernah teridentifikasi sebelumnya.

B. Rumusan Masalah

1. Bagaimana membangun model *deep learning* untuk mengklasifikasikan serangan siber?
2. Bagaimana menentukan algoritma *deep learning* terbaik untuk mengklasifikasikan serangan siber?

C. Batasan Penelitian

1. Penelitian ini hanya akan menggunakan dataset yang dibuat sendiri, yang diperoleh dari hasil rekaman lalu lintas jaringan secara langsung. Dataset ini diambil dan dikumpulkan melalui proses simulasi serangan yang dilakukan dalam lingkup virtual yang dimana Kali Linux sebagai penyerang dan Server Ubuntu sebagai korban. Seluruh aktivitas jaringan direkam di sisi Server Ubuntu menggunakan Wireshark. Dataset ini dirancang untuk menggambarkan serangan siber secara realistis yang sesuai dengan kebutuhan penelitian
2. Penelitian ini akan fokus pada serangan siber yang umum seperti *DoS*, *Port Scanning*, *Brute Force Attack*, dan *Spoofing*. Serangan siber lain yang kurang umum tidak akan dianalisis dalam penelitian ini.
3. Hanya beberapa algoritma *deep learning* yang akan digunakan dan dibandingkan, seperti *Deep Neural Network* (DNN), *Bidirectional Gated Recurrent Unit* (BiGRU) dan *Long Short-Term Memory* (LSTM). Algoritma lain yang lebih kompleks atau tidak relevan tidak akan dibahas secara mendalam.

D. Tujuan Penelitian

1. Mengembangkan model *deep learning* untuk mengklasifikasikan serangan siber.
2. Menentukan algoritma *deep learning* yang paling tepat untuk mengklasifikasikan serangan siber melalui proses analisis dan perbandingan performa beberapa algoritma berdasarkan data lalu lintas jaringan.

E. Metode Penelitian

Metode penelitian yang akan digunakan dalam pengembangan sistem “Klasifikasi Intrusi Jaringan Menggunakan Algoritma *Deep Learning*” adalah sebagai berikut:

1. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan eksperimen kuantitatif dengan penerapan algoritma *deep learning* untuk mendeteksi intrusi jaringan. Dataset merupakan hasil dari rekaman lalu lintas jaringan menggunakan Wireshark yang terpasang pada Server Ubuntu (target), dan penyerang Kali Linux. Semua simulasi serangan dilakukan pada lingkup *virtual*, dan menggunakan jaringan pribadi (*Tetring*). Hasil dari rekaman lalu lintas jaringan ini yang nantinya digunakan sebagai Dataset yang dimana sebagai sumber data utama, karena mencakup berbagai serangan siber modern yang relevan, seperti DoS, Brute Force Attack, Port Scanning dan Spoofing.

2. Tahapan Penelitian

Penelitian dilakukan melalui tahapan berikut:

a. Studi Literatur

Penelitian ini diawali dengan mempelajari teori-teori yang berkaitan dengan intrusi jaringan serta pendekatan algoritma *deep learning* yang relevan, seperti klasifikasi, dan pengenalan pola. Setelah pemahaman teoritis diperoleh, langkah selanjutnya adalah mengkaji karakteristik dataset hasil rekaman jaringan sendiri yang digunakan sebagai data latih dan uji dalam penelitian. Dataset ini dikumpulkan melalui simulasi lalu lintas jaringan nyata yang mencakup aktivitas normal dan beragam serangan siber umum, kemudian direkam menggunakan Wireshark, diolah menjadi aliran data menggunakan CICFlowMeter. Analisis dataset ini meliputi struktur atau format data, jenis-jenis fitur yang tersedia, serta kategori serangan yang berhasil direkam. Proses ini bertujuan untuk memastikan kesesuaian dataset dengan kebutuhan

pengembangan model *deep learning*, sekaligus memahami distribusi dan representasi data agar dapat dilakukan proses pra-proses, pelabelan, dan pelatihan model secara optimal. Kombinasi antara kajian teoritis dan analisis dataset diharapkan dapat menjadi dasar yang kuat dalam merancang dan mengimplementasikan IDS jaringan berbasis *deep learning* yang akurat dan adaptif terhadap berbagai serangan siber.

b. Pengumpulan dan Pemrosesan Dataset

Pada tahap persiapan data ini, dataset ini diambil dari hasil rekaman lalu lintas jaringan menggunakan Wireshark yang terinstal pada Server Ubuntu. Server Ubuntu ini berperan sebagai korban, sedangkan Kali Linux berperan sebagai penyerang yang nantinya akan menghasilkan lalu lintas jaringan yang tidak wajar, lalu akan *capture* menggunakan aplikasi Wireshark. Semua proses perekaman lalu lintas jaringan ini dilakukan pada ruang lingkup *virtual*. Proses perekaman ini dilakukan beberapa kali, sehingga mendapatkan dataset yang dibutuhkan untuk keperluan penelitian. Dataset kemudian akan digunakan sebagai sumber data utama untuk pelatihan dan pengujian model deteksi intrusi jaringan yang dikembangkan. Pemilihan metode perekaman data yang independen ini memungkinkan kontrol penuh terhadap serangan siber dan kondisi jaringan yang akan disimulasikan, sehingga representasi penelusuran jaringan lebih mendekati skenario yang diharapkan.

Setelah proses perekaman selesai, maka akan dilanjutkan ke tahap pembersihan data. Langkah ini sangat penting untuk mengatasi berbagai masalah kualitas data, seperti nilai yang hilang, duplikasi atau ketidakkonsistenan yang mungkin muncul pada saat perekaman. Proses pembersihan ini bertujuan untuk menghasilkan data bersih yang siap digunakan dalam analisis lebih lanjut, memastikan bahwa model tidak belajar dari data yang cacat, yang dapat mempengaruhi hasil dari penelitian ini.

Selanjutnya, normalisasi data dilakukan. Normalisasi ini berfungsi untuk menskalakan nilai setiap fitur dalam dataset agar sesuai dengan kebutuhan input model *deep learning* yang dikembangkan. Normalisasi sangat penting dilakukan agar algoritma yang dikembangkan, dapat belajar dengan optimal, menghindari dominasi fitur tertentu yang memiliki nilai yang tidak konsisten, dan mempercepat konvergensi pada saat pelatihan model. Hal ini akan berkontribusi pada efisiensi dan akurasi model yang dibangun.

Setelah data dinormalisasi, dilakukan eksplorasi data untuk memahami distribusi dan proporsi lalu lintas jaringan, baik yang tergolong normal maupun yang mengandung serangan. Hasil dari eksplorasi ini kemudian digunakan dalam tahap ekstraksi fitur, di mana hanya fitur-fitur penting seperti alamat IP, nomor *port*, jenis protokol, durasi koneksi, jumlah paket, dan *byte* yang dikirim yang akan dipilih untuk dijadikan input model. Setelah semua tahapan ini selesai, dataset kemudian dibagi menjadi tiga bagian, yaitu data latih, data validasi, dan data uji, guna memastikan proses pelatihan dan evaluasi model dilakukan secara sistematis dan akurat.

- 1) Eksplorasi data: Memahami distribusi data, proporsi lalu lintas normal dan serangan, serta fitur yang relevan.
- 2) Ekstraksi fitur: Memilih fitur penting untuk deteksi intrusi, seperti:
 - a) *Source/Destination* IP dan Port
 - b) Jenis protokol
 - c) Waktu koneksi
 - d) Jumlah paket dan *byte* yang dikirim.

c. Perancangan Model

Setelah algoritma *deep learning* yang akan dibandingkan (DNN, LSTM, dan BiGRU) dipilih, langkah selanjutnya dalam tahap pengembangan model adalah merancang arsitektur dari masing-masing model tersebut. Perancangan ini dilakukan menggunakan *framework*

TensorFlow untuk memastikan implementasi yang konsisten dan efisien.

Dalam perancangan arsitektur, beberapa komponen kunci akan ditentukan:

- 1) Jumlah Lapisan (*Number of Layers*): Model akan dirancang dengan menentukan jumlah lapisan *hidden* yang optimal. Penentuan ini akan mempertimbangkan kompleksitas data dan kebutuhan untuk menangkap pola-pola rumit.
- 2) Penerapan *Dropout*: Teknik *dropout* akan diterapkan pada lapisan-lapisan tertentu untuk mencegah *overfitting*. *Dropout* membantu meningkatkan kemampuan generalisasi model dengan secara acak mengabaikan sejumlah *neuron* selama pelatihan.
- 3) Fungsi Aktivasi (*Activation Functions*): Fungsi aktivasi seperti ReLU (*Rectified Linear Unit*) atau *Sigmoid* akan dipilih untuk setiap lapisan. Pemilihan fungsi aktivasi ini krusial dalam memperkenalkan *non-linearitas* ke dalam model, yang memungkinkan model mempelajari hubungan yang lebih kompleks dalam data.
- 4) Algoritma Optimasi (*Optimization Algorithm*): Algoritma optimasi seperti Adam atau RMSprop akan digunakan untuk mempercepat proses konvergensi model selama pelatihan. Algoritma ini bertanggung jawab untuk menyesuaikan bobot model berdasarkan gradien *loss function*.

Selain arsitektur dasar, penentuan *hiperparameter* juga merupakan aspek penting dalam perancangan model. *Hiperparameter* ini, seperti ukuran *batch* (*batch size*), jumlah epoch (*number of epochs*), dan tingkat pembelajaran (*learning rate*), akan diatur secara cermat. Pemilihan *hiperparameter* yang tepat akan sangat memengaruhi akurasi dan kinerja akhir dari IDS berbasis *deep learning* yang dikembangkan. Proses perancangan ini dilakukan secara iteratif, di mana berbagai

konfigurasi arsitektur dan *hiperparameter* dapat diuji untuk menemukan kombinasi terbaik yang menghasilkan akurasi tertinggi dalam deteksi intrusi jaringan.

Setelah model selesai dilatih, langkah berikutnya adalah menguji performanya menggunakan data uji yang telah dipisahkan di awal. Pengujian ini bertujuan untuk menilai kemampuan model dalam mendeteksi serangan jaringan secara akurat pada data yang belum pernah dilihat sebelumnya.

Untuk mengukur performa tersebut, berbagai metrik evaluasi digunakan untuk memberikan gambaran menyeluruh tentang efektivitas model:

- 1) Akurasi: Mengukur persentase total prediksi yang benar terhadap keseluruhan data uji.
- 2) Presisi (*Precision*): Menunjukkan seberapa banyak dari prediksi serangan yang benar-benar merupakan serangan.
- 3) Sensitivitas (*Recall*): Mengukur sejauh mana model berhasil mendeteksi seluruh serangan yang ada dalam data uji.
- 4) *False Positive Rate* (FPR): Menghitung seberapa sering model salah mendeteksi lalu lintas normal sebagai serangan. Metrik ini sangat penting dalam konteks sistem keamanan jaringan karena *false positive* yang tinggi dapat menyebabkan peringatan palsu yang berlebihan.

Evaluasi komprehensif menggunakan metrik-metrik ini akan memberikan gambaran yang akurat tentang keandalan dan efektivitas model dalam mendeteksi intrusi jaringan secara nyata.

d. Implementasi

Setelah didapatkan model dikembangkan dan mendapatkan model terbaik, tahap selanjutnya adalah implementasi model tersebut ke dalam antarmuka pengguna berbasis web. Model yang telah dilatih diekspor menjadi *file* .keras agar dapat langsung di-*import* ke dalam sistem *website* yang dibangun menggunakan *framework* Streamlit versi

1.45.1. Streamlit dipilih karena kemudahannya dalam membuat aplikasi *web* interaktif langsung dari *script* Python, serta integrasinya yang kuat dengan pustaka *data science* populer, fitur visualisasi *real-time*, dan *deployment* yang efisien. Dalam aplikasi *web* ini, pengguna dapat mengunggah berkas rekaman jaringan dalam format CSV, yang kemudian akan melalui serangkaian proses pra-pemrosesan data esensial, seperti pembersihan dan konversi tipe data oleh fungsi `clean_and_convert_data()`, sebelum data diumpankan ke model *deep learning* untuk deteksi intrusi.

3. Alat yang Digunakan

a. Untuk Simulasi Serangan:

1) Perangkat Keras:

- a) Prosesor AMD Ryzen 7
- b) Kartu Grafis NVIDIA RTX 4050 6 GB
- c) RAM 16 GB

2) Perangkat Lunak

- a) Windows 11 Home
- b) VirtualBox-7.1.8
- c) Ubuntu-24.04.2
- d) Kali-Linux-2025

b. Untuk Pengembangan Model

1) Perangkat Lunak

- 1) Sistem operasi: Ubuntu 22.04 LTS (64-bit)
- 2) Pustaka *deep learning*: TensorFlow
- 3) Alat analisis data: Pandas, NumPy, dan Matplotlib

2) Perangkat Keras:

- a) Server Case: Supermicro SYS-621C-TN12R - X13 CloudDC 2U dengan motherboard X13DDW-A
- b) Prosesor: $2 \times$ Intel Xeon EMR 4510 (12 core, 2.4 GHz, 30MB cache, TDP 150W)
- c) RAM: 128 GB DDR5 ECC REG (4×32 GB DDR5-4800)
- d) GPU: $2 \times$ NVIDIA Ada L40S 48GB GDDR6 PCIe Gen 4
- e) Penyimpanan: Mendukung hingga $12 \times 3.5''$ drive (detail tergantung konfigurasi).

4. Dataset

Dataset yang digunakan merupakan dataset sintetik hasil simulasi serangan jaringan yang dikumpulkan secara mandiri menggunakan Teknik *penetration testing*.

F. Sistematika Penulisan

Sistematika penulisan skripsi ini disusun secara sistematis agar memudahkan pembaca dalam memahami alur penelitian yang dilakukan. Adapun sistematika penulisan dalam skripsi ini terdiri dari 7 bab, yaitu sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan tugas akhir.

BAB II TINJAUAN PUSTAKA

Bab ini berisi tinjauan pustaka yang membahas berbagai penelitian terdahulu terkait penerapan *deep learning* dan *machine learning* dalam IDS jaringan. Selain itu, bab ini juga menguraikan konsep dasar keamanan siber, serta menyajikan tabel perbandingan dari beberapa penelitian

untuk memberikan gambaran komprehensif mengenai pendekatan yang telah digunakan dalam mendeteksi ancaman siber.

BAB III LANDASAN TEORI

Landasan teori mencakup teori-teori terdiri atas: Serangan Siber, *Deep Learning*, dan *Intrusion Detection System (IDS)*

BAB IV PENGUMPULAN DATASET

Bab ini berisi tentang penjelasan mengenai alur yang dilakukan penulis dalam proses pengumpulan dataset yang digunakan dalam penelitian ini. Mulai dari perekaman lalu lintas jaringan, sampai menjadi dataset yang dapat digunakan untuk pengembangan model *deep learning*.

BAB V PENGEMBANGAN MODEL

Bab ini berisi tentang penguraian tahapan pengembangan model *deep learning* untuk deteksi intrusi jaringan, dimulai dari pemuatan dan pra-pemrosesan akhir data, perancangan arsitektur, untuk setiap model (DNN, LSTM, BiGRU), konfigurasi parameter pelatihan, hingga proses evaluasi komprehensif menggunakan metrik kinerja dan *Confusion Matrix*

BAB VI IMPLEMENTASI MODEL

Bab ini berisi tentang penjelasan secara rinci implementasi model *deep learning* yang telah dilatih ke dalam sebuah *Intrusion Detection System (IDS)* berbasis web. Bagian ini mencakup pemilihan model terbaik berdasarkan evaluasi, integrasi model ke dalam antarmuka pengguna, serta proses pengujian fungsionalitas secara keseluruhan.

BAB VII PENUTUP

Bab ini berisi tentang kesimpulan dari keseluruhan penelitian yang telah dilakukan mengenai deteksi intrusi jaringan menggunakan algoritma *deep learning* serta

memberikan saran untuk pengembangan di masa mendatang.

