Laporan Tugas Besar Kelompok Dasar Kecerdasan Artifisial (DKA)
“Network Intrusion Detection”

Disusun oleh:

MUHAMMAD IRGIANSYAH	103012300039
BILL STEPHEN SEMBIRING	103012330197

Program Studi S1 Informatika

Fakultas Informatika

Universitas Telkom

Bandung

2025

Deskripsi Dataset: Network Intrusion Detection
https://www.kaggle.com/datasets/sampadab17/network-intrusion-detection

Latar Belakang

Seiring pesatnya perkembangan komputer, aktivitas komunikasi dan pertukaran data melalui jaringan komputer menjadi hal umum dalam kehidupan sehari-hari.Namun dalam kemajuan teknologi, Dalam bidang keamanan jaringan, deteksi intrusi merupakan elemen vital untuk menjaga integritas sistem informasi dari berbagai ancaman eksternal. Meningkatnya aktivitas siber yang berpotensi merusak menuntut solusi cerdas untuk mengidentifikasi lalu lintas yang mencurigakan secara otomatis. Oleh karena itu, pendekatan berbasis data, seperti klasifikasi menggunakan machine learning, menjadi semakin penting. Dataset ini memberikan landasan untuk membangun sistem pendeteksi intrusi yang mampu membedakan antara aktivitas normal dan serangan berbahaya.

Rumusan Masalah

1. Dataset yang digunakan merupakan data simulasi dan tidak mencakup data jaringan.
2. Klasifikasi dilakukan dalam dua pendekatan:
    - Biner (*normal* vs *intrusion*).
    - Multi-kelas (DoS, Probe, R2L, U2R).
3. Algoritma yang digunakan dibatasi pada dua pendekatan utama:
    - **K-Nearest Neighbors (KNN)**: Metode klasifikasi berbasis kedekatan fitur antara data uji dan data latih.
    - **Fuzzy Logic**: Digunakan untuk mengakomodasi ketidakpastian dan ambiguitas dalam data, terutama dalam menentukan kelas berdasarkan aturan logika fuzzy mamdani dan sugeno.
4. Evaluasi hanya dilakukan menggunakan data uji dari dataset yang sama, tanpa implementasi pada jaringan.
5. Proyek tidak mencakup integrasi model ke dalam sistem IDS secara operasional.

Dataset

Dataset ini merupakan himpunan data simulasi lalu lintas jaringan yang dikumpulkan untuk mendeteksi aktivitas intrusi. Dataset ini mencakup ribuan sampel koneksi dengan berbagai fitur yang merepresentasikan karakteristik teknis dari masing-masing koneksi, seperti durasi, protokol, jumlah paket, dan atribut perilaku lainnya. Label data menunjukkan apakah sebuah koneksi tergolong normal atau merupakan serangan (misalnya DoS, probing, R2L, U2R).

Deskripsi Fitur

Beberapa fitur penting yang tersedia dalam dataset meliputi:

- **Duration**: Lamanya koneksi berlangsung.
- **Protocol_type**: Jenis protokol yang digunakan (TCP, UDP, ICMP).
- **Service**: Layanan jaringan yang digunakan (http, ftp, smtp, dll).
- **Flag**: Status koneksi berdasarkan protokol.
- **Src_bytes & Dst_bytes**: Jumlah byte yang dikirim dari dan ke sumber.
- **Count & srv_count**: Jumlah koneksi ke host atau layanan dalam jangka waktu tertentu.

Label target berupa klasifikasi biner (normal/intrusion) atau multi-kelas (tipe-tipe serangan).

Permasalahan

Permasalahan utama adalah mengembangkan model klasifikasi yang dapat secara akurat mengidentifikasi apakah sebuah koneksi jaringan merupakan aktivitas normal atau bentuk serangan. Sistem ini dapat digunakan untuk meningkatkan efisiensi tim keamanan siber dalam melakukan deteksi dini terhadap ancaman.

Tujuan

- Menerapkan algoritma pembelajaran mesin untuk mendeteksi intrusi jaringan.
- Mengevaluasi performa model menggunakan metrik seperti akurasi, precision, recall, dan F1-score.
- Menganalisis efektivitas fitur dalam mengidentifikasi berbagai jenis serangan.

Alasan Pemilihan Dataset

- Relevan dengan isu terkini dalam bidang keamanan jaringan.
- Struktur data yang kaya dan cocok untuk pendekatan klasifikasi.
- Cocok digunakan untuk riset dan pengembangan sistem IDS (Intrusion Detection System).
- Dilengkapi dengan label dan fitur yang memadai untuk pemodelan dan evaluasi performa.

Metodologi

1. Pengumpulan dan Eksplorasi Dataset

- Memahami struktur dataset, fitur-fitur yang tersedia, dan distribusi label.

2. Pra-pemrosesan Data

- Encoding fitur kategorikal (misalnya protocol_type, service, flag).
- Normalisasi data numerik untuk mendukung performa KNN.
- Penanganan data yang hilang, duplikat, atau tidak konsisten.

3. Pembagian Dataset

- Dataset dibagi menjadi data latih dan data uji (misalnya 80:20 split).

4. Implementasi Algoritma

- **KNN**:
    - Menentukan nilai **k** yang optimal.
    - Menghitung jarak (misalnya Euclidean) antara data uji dan data latih.
    - Klasifikasi berdasarkan mayoritas tetangga terdekat.
- **Fuzzy Logic**:
    - Menentukan **fuzzy sets** untuk setiap fitur (contoh: *rendah*, *sedang*, *tinggi*).
    - Menetapkan aturan fuzzy (fuzzy rules) berdasarkan pengetahuan domain.
    - Melakukan proses inferensi dan defuzzifikasi untuk klasifikasi.

5. Evaluasi Model

- Menggunakan metrik:
    - Akurasi
    - Precision

    - Recall
    - F1-score
- Perbandingan performa antara KNN dan Fuzzy.

6. block diagram pada fuzzy dan knn

7. Analisis

- Analisis kekuatan dan kelemahan dari kedua metode.
- Identifikasi fitur yang paling berpengaruh terhadap performa model.

8. Penyusunan Laporan

- Menyusun hasil eksperimen dan visualisasi performa model.
- Diskusi dan kesimpulan.