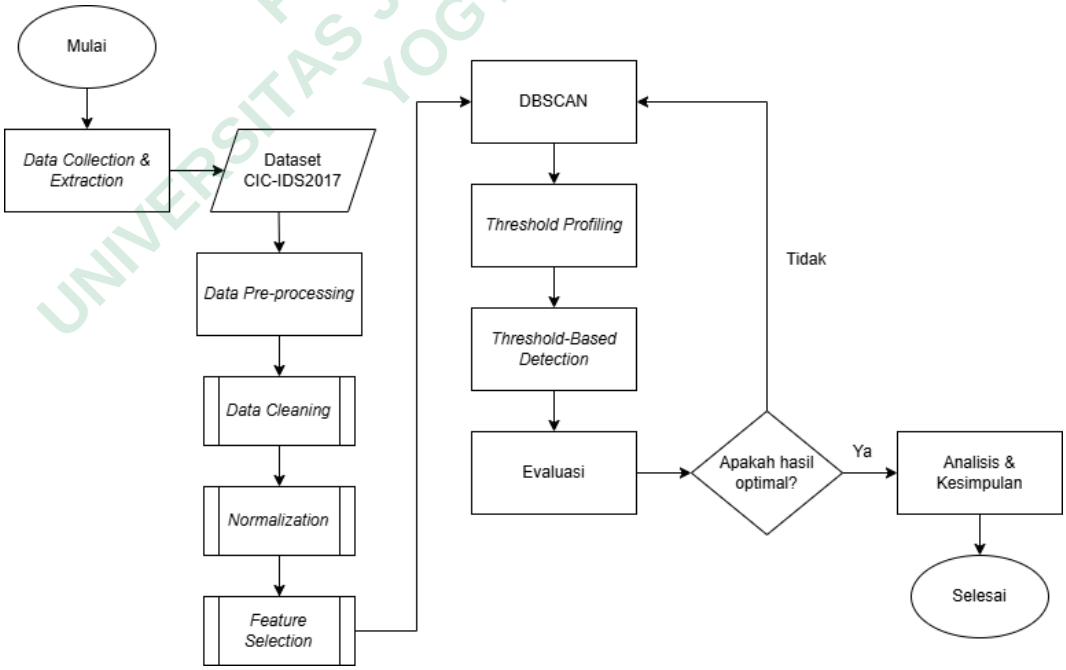
BAB 3
METODE PENELITIAN

Penelitian ini bertujuan untuk mengimplementasikan metode threshold profiling berbasis algoritma DBSCAN dalam mendeteksi anomali pada lalu lintas jaringan serangan DDoS. Pendekatan yang digunakan adalah Threshold-Based Detection, dimana proses menentukan nilai ambang batas (threshold profiling) berdasarkan hasil *clustering*. Penelitian diawali dengan identifikasi permasalahan pada deteksi anomali oleh IDS terkait tingginya *false alarm*. Selanjutnya, dilakukan serangkaian eksperimen *unsupervised learning* menggunakan DBSCAN untuk membentuk *cluster* dengan menggunakan dataset CIC-IDS2017 sebagai bahan pengujian. Sebagai tahap evaluasi, dengan mengukur efektivitas deteksi terhadap lalu lintas jaringan menggunakan metrik evaluasi yang relevan seperti *Accuracy, Precision*, *Recall*, *F1-Score*, dan FPR. Gambar 3.1 menyajikan alur lengkap dari tahapan metode penelitian yang dilakukan.



Gambar 3. 1 Alur Penelitian

3.1 Bahan dan Alat Penelitian

Bahan yang digunakan dalam penelitian ini adalah dataset CIC-IDS2017 yang mencakup lalu lintas jaringan normal dan anomali jenis serangan DDoS. Alat yang digunakan dalam penelitian ini adalah komputer dengan spesifikasi yang memadai untuk menjalankan pemrosesan data, threshold profiling, dan DBSCAN. Adapun sistem operasi, perangkat keras, dan perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Sistem Operasi: Windows 10 atau lebih baru
2. Perangkat keras:
    - Prosesor: Intel Core i3 atau lebih tinggi
    - RAM: Minimal 4GB
    - Penyimpanan: SSD dengan kapasitas min 256 GB atau lebih untuk meningkatkan kecepatan akses data
3. Perangkat lunak:
    - Bahasa pemrograman Python dengan *library* Scikit-Learn, Pandas, NumPy, Matplotlib, dan lain-lain
    - Google Colab

3.2 Jalan Penelitian

Penelitian ini menggunakan pendekatan eksperimen dan analisis data, dengan alur metode yang ditunjukkan pada Gambar 3.1. Tahapan-tahapan penelitian yang dijelaskan sebagai berikut:

1. Tahap Awal
    a. Mengkaji permasalahan pada deteksi anomali jaringan menggunakan IDS pada aspek tingkat *false alarm* yang tinggi.
    b. Mengusulkan solusi berbasis Threshold-Based Detection dengan algoritma DBSCAN untuk mendeteksi anomali lalu lintas jaringan serangan DDoS.
2. *Data Collection & Extraction*
    a. Menggunakan dataset CIC-IDS2017, khususnya serangan DDoS.

b. Data dikumpulkan dalam format *csv* untuk memudahkan pemrosesan.

3. *Data Pre-processing*
   a. *Data cleaning* (pembersihan data) untuk menghapus duplikasi, *missing value*, dan fitur yang tidak relevan.
   b. *Normalization* (normalisasi) untuk menyamakan skala nilai fitur menggunakan *Min-Max Scaling*.
   c. *Feature Selection* (seleksi fitur) untuk meningkatkan akurasi deteksi dan mengurangi data yang tidak perlu. Setelah *feature selection*, dilakukan *data splitting* (pembagian data) menjadi data training (benign) dan data testing (benign dan DDoS) dengan rasio 70:30 untuk memisahkan proses pembentukan threshold dari proses evaluasi.
4. DBSCAN
   a. DBSCAN digunakan untuk mendeteksi anomali berdasarkan kepadatan data dengan mengidentifikasi anomali (*noise*) yang tidak masuk dalam *cluster* manapun.
   b. Dengan parameter utama:
      - *Epsilon* (ε): Jarak maksimal antar titik dalam satu klaster.
      - *MinPts*: Jumlah minimal titik dalam radius ε untuk membentuk *cluster*.
5. Threshold Profiling
   a. Threshold profiling digunakan untuk membedakan lalu lintas normal dan anomali berdasarkan fitur tertentu.
   b. Threshold dihitung berdasarkan distribusi nilai dari lalu lintas normal. Perhitungan threshold menggunakan pendekatan statistik berdasarkan penjumlahan rata-rata dan standar deviasi dari fitur yang dianalisis.
   c. Hasil klasterisasi DBSCAN digunakan sebagai dasar pembentukan threshold, dengan menggunakan data dari klaster valid yang merepresentasikan lalu lintas benign. Rumus penghitungan tersebut dapat dilihat pada Persamaan 1 (Sukma Aji dkk., 2024).

$$T = \mu + k \, . \sigma \quad (1)$$

Keterangan:

T: Nilai threshold (ambang batas)

μ: Rata-rata

k: Koefisien pengali

σ: Standar deviasi

6. Threshold-Based Detection
   a. Setelah threshold dihitung, proses klasifikasi menggunakan rumus yang dapat dilihat pada Persamaan 2.
   b. Penelitian yang sama oleh Sukma Aji dkk., (2024) ini digunakan untuk mendeteksi serangan DDoS berdasarkan jika satu atau lebih fitur dari suatu aliran jaringan melebihi threshold, maka diklasifikasikan sebagai anomali.

$$g\,(x) = \begin{cases} 1, x \geq T \\ 0, x < T \end{cases} \quad (2)$$

Keterangan:

$g\,(x)$: Hasil klasifikasi

x: Nilai aktual dari fitur yang dianalisis

T: Threshold dari Persamaan 1

7. Evaluasi
   a. Data testing diseimbangkan terlebih dahulu (*data balancing*) melalui teknik *undersampling* pada distribusi label benign dan DDoS guna mencegah bias evaluasi akibat salah satu kelas yang lebih dominan.
   b. Evaluasi dilakukan terhadap hasil deteksi anomali dengan menggunakan metrik evaluasi *Confusion Matrix*. Metrik ini digunakan untuk menggambarkan kinerja suatu sistem dengan membandingkan hasil prediksi terhadap label sebenarnya (Kartyasa dkk., 2015).
      - *Accuracy*: presentase dari prediksi benar dari seluruh prediksi yang dilakukan model, dengan rumus yang ditunjukkan pada Persamaan 3.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

- *Precision*: ukuran estimasi probabilitas prediksi positif yang benar, dengan rumus yang ditunjukkan pada Persamaan 4.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

- *Recall*: persentase dari prediksi positif yang benar dari seluruh data positif yang ada, dengan rumus yang ditunjukkan pada persamaan 5.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

- *F1-Score*: rata-rata dari *Precision* dan *Recall*, dengan rumus yang ditunjukkan pada Persamaan 6.

$$F1\text{-}Score = 2 \text{ x } \frac{Precision\ x\ Recall}{Precision + Recall} \quad (6)$$

- *False Positive Rate* (FPR): rasio prediksi salah positif terhadap seluruh data negatif yang sebenarnya. Dalam konteks penelitian ini, FPR mengukur data normal yang salah terdeteksi sebagai serangan DDoS. FPR dihitung menggunakan rumus yang ditunjukkan pada Persamaan 7.

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

Keterangan:
FP: *False Positive*
FN: *False Negative*
TP: *True Positive*

TN: *True Negative*

- *Silhouette Index* (SI) atau *Silhouette Score* untuk evaluasi *clustering* DBSCAN, dengan rumus yang ditunjukkan pada Persamaan 8 (Ibadirachman dkk., 2024).

$$\text{SI} = \frac{b - \alpha}{max\ (\alpha,b)} \qquad (8)$$

Keterangan:

α: jarak rata-rata intra-klaster

b: jarak rata-rata ke klaster terdekat

- Jika hasil evaluasi belum memuaskan, proses kembali ke tahap DBSCAN dan/atau threshold profiling untuk penyesuaian parameter, jika hasil sudah optimal proses dilanjutkan ke tahap analisis & kesimpulan.

8. Analisis & Kesimpulan
   a. Menganalisis efektivitas threshold profiling dengan algoritma DBSCAN dalam mendeteksi anomali.
   b. Menyimpulkan seberapa baik metode ini mampu meningkatkan akurasi dan mereduksi *false alarm.*
   c. Kesimpulan berdasarkan evaluasi metrik dan visualisasi.