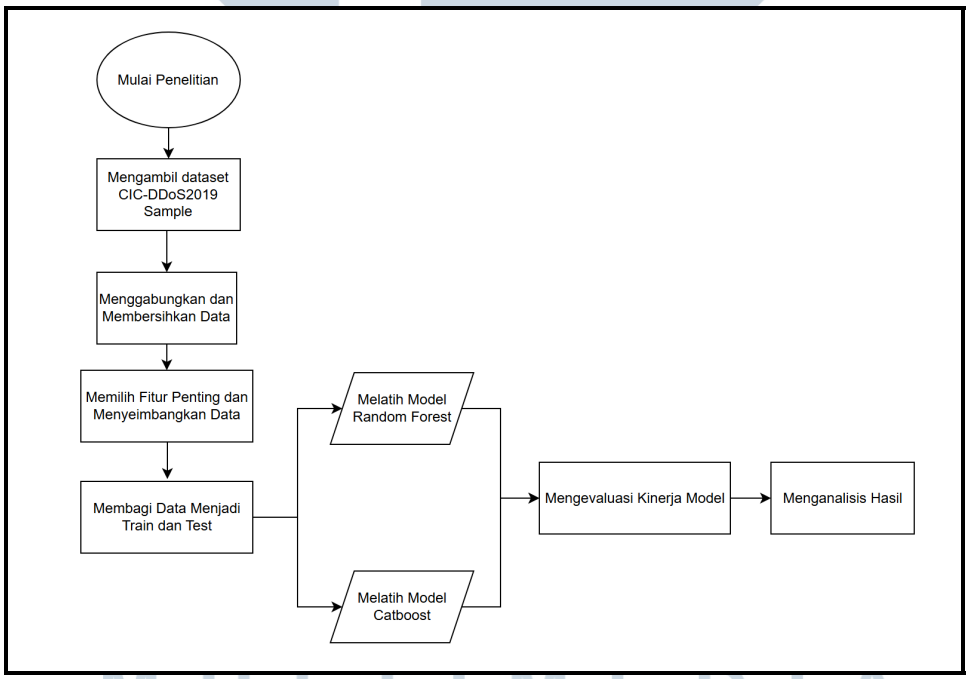
BAB 3
METODOLOGI PENELITIAN

3.1 Jenis dan Pendekatan Penelitian

Penelitian ini merupakan jenis penelitian kuantitatif dengan pendekatan eksperimen komputasional. Tujuan dari penelitian ini adalah untuk menguji dan membandingkan kinerja algoritma *Random Forest* dan *CatBoost* dalam mendeteksi serangan Distributed Denial of Service (DDoS) menggunakan dataset CIC-DDoS2019, dengan evaluasi berbasis metrik *precision*, *recall*, dan *F1-score*.

3.2 Alur Penelitian

Penelitian ini dilakukan melalui tahapan yang terstruktur sebagaimana ditunjukkan pada Gambar 3.1:



Gambar 3.1. Diagram alur penelitian

1. **Mengambil Dataset CIC-DDoS2019 (Sample)**

   Penelitian ini dimulai dengan pengambilan data dari dataset CIC-DDoS2019 versi sampel, yang terdiri atas lima file yaitu: `LDAP_sample.csv`, `MSSQL_sample.csv`, `NetBIOS_sample.csv`, `Portmap_sample.csv`, dan

`Syn_sample.csv`. Masing-masing file berisi data lalu lintas jaringan termasuk serangan DDoS dan trafik normal (benign).

2. **Menggabungkan dan Membersihkan Data**
   Seluruh file digabungkan menjadi satu *dataframe* utama. Kemudian dilakukan pembersihan data dengan menghapus kolom yang tidak relevan seperti `Flow ID`, `Source IP`, `Destination IP`, dan `Timestamp`. Selain itu, data yang mengandung nilai `NaN`, `inf`, atau `-inf` ditangani dengan metode imputasi median.

3. **Memilih Fitur Penting dan Menyeimbangkan Data**
   Dari seluruh fitur yang tersedia, dipilih sejumlah fitur penting berdasarkan analisis domain dan korelasi terhadap target klasifikasi. Selanjutnya, dilakukan proses penyeimbangan (*balancing*) dataset agar tiap kelas (termasuk benign dan masing-masing jenis serangan) memiliki jumlah data yang setara. Teknik yang digunakan adalah *undersampling* pada tiap kelas sebanyak 200 data.

4. **Membagi Data Menjadi Data Latih dan Data Uji**
   Data yang telah dibersihkan dan diseimbangkan dibagi menjadi data latih (70%) dan data uji (30%) menggunakan teknik *stratified split*. Pendekatan ini menjaga distribusi proporsional antar kelas pada kedua subset agar tidak terjadi bias selama pelatihan maupun evaluasi.

5. **Melatih Model Random Forest dengan Tuning Otomatis**
   Model Random Forest dilatih menggunakan data latih multiclass dengan pemilihan parameter optimal melalui teknik *Grid Search*. Parameter seperti `n_estimators`, `max_depth`, dan `min_samples_split` dievaluasi secara sistematis untuk mendapatkan kombinasi terbaik berdasarkan skor F1 makro. Model ini dipilih karena kestabilannya dan kemampuannya dalam menangani fitur numerik.

6. **Melatih Model CatBoost dengan Tuning Otomatis**
   Model CatBoost dilatih menggunakan pendekatan *Randomized Search* untuk menyempurnakan parameter `iterations`, `depth`, dan `learning_rate`. CatBoost dikenal efisien dalam menangani data tabular dan mampu mengelola distribusi data kompleks tanpa memerlukan one-hot encoding.

7. **Mengevaluasi Kinerja Model**
   Evaluasi performa dilakukan menggunakan metrik `accuracy`, `precision`, `recall`, dan `F1-score` untuk masing-masing kelas dan dalam bentuk rata-rata makro. Evaluasi dilakukan pada data uji dan diperkuat dengan teknik *5-fold cross-validation* untuk mengukur konsistensi model. Visualisasi hasil disajikan melalui *heatmap*, *line chart*, dan *confusion matrix*.

8. **Menganalisis Hasil Evaluasi**
   Hasil dari kedua model dibandingkan dan dianalisis secara mendalam, baik dari segi akurasi, waktu pelatihan, maupun stabilitas prediksi. Analisis juga mempertimbangkan perbedaan algoritmik antara Random Forest dan CatBoost serta relevansi penggunaannya dalam sistem deteksi serangan DDoS pada infrastruktur jaringan enterprise.

Penjelasan tersebut mencerminkan secara komprehensif setiap tahapan yang dilakukan dalam penelitian ini, sesuai dengan alur praktik yang diimplementasikan pada Jupyter Notebook yang telah digunakan peneliti.

3.3 Tools dan Lingkungan Pengujian

Eksperimen dilakukan menggunakan:

- Bahasa pemrograman: Python 3.10
- Platform: Jupyter Notebook
- Dataset: CIC-DDoS2019 (dalam format .csv, diunduh dari Canadian Institute for Cybersecurity)