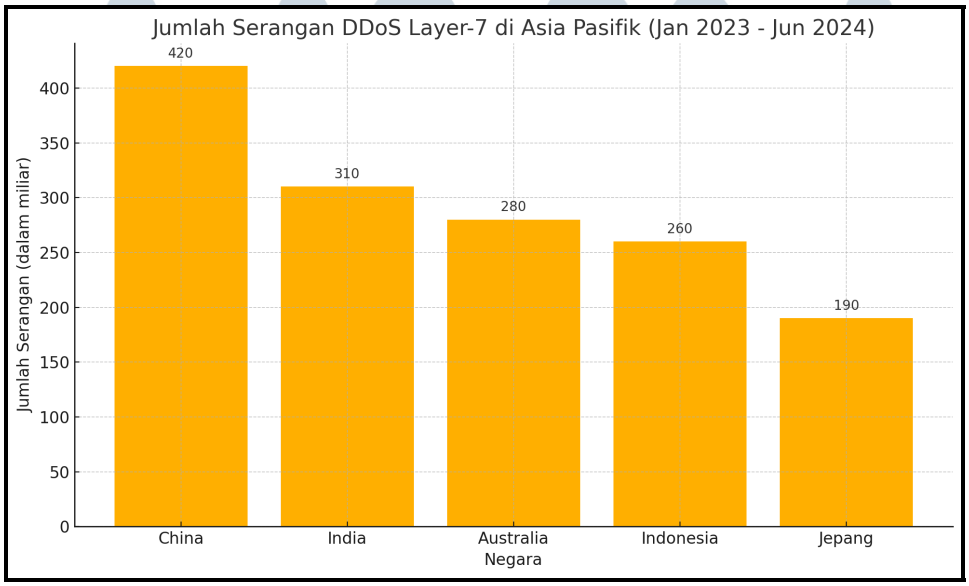
BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Pada era transformasi digital saat ini, serangan siber Distributed Denial of Service (DDoS) menjadi ancaman yang semakin mengkhawatirkan, termasuk di Indonesia. Serangan DDoS berusaha membuat layanan atau jaringan tidak dapat diakses pengguna dengan membanjiri traffic palsu dalam jumlah besar ke target, sehingga menghabiskan sumber daya dan melumpuhkan layanan tersebut. Tren serangan DDoS menunjukkan peningkatan signifikan secara global maupun lokal. Di Indonesia sendiri, jumlah serangan DDoS dilaporkan terus meningkat sejak awal 2023. Menurut data terbaru dari Akamai, sepanjang Januari 2023 hingga Juni 2024, Indonesia menerima total 260 miliar serangan DDoS pada lapisan aplikasi (layer-7) – menempatkan Indonesia di peringkat ke-4 negara paling banyak mendapat serangan DDoS di kawasan Asia Pasifik [1] [2]. Secara regional, aktivitas DDoS di Asia Pasifik dilaporkan naik hingga 5 kali lipat dalam kurun 18 bulan terakhir [1]. Angka-angka ini mengindikasikan betapa serius dan aktualnya ancaman DDoS, sehingga topik ini dipilih sebagai fokus penelitian karena relevan dengan kondisi nyata yang dihadapi berbagai institusi digital di Indonesia saat ini.



Gambar 1.1. Distribusi serangan DDoS secara global dan posisi Indonesia

Ancaman DDoS terhadap infrastruktur jaringan enterprise menimbulkan

permasalahan krusial. Serangan DDoS dapat berlangsung dari hitungan jam hingga berhari-hari, menyebabkan downtime layanan, mengganggu operasional bisnis, dan berpotensi menimbulkan kerugian besar. Secara langsung, serangan DDoS mengakibatkan performa sistem menurun drastis atau bahkan tidak dapat diakses sama sekali oleh pengguna yang sah [3]. Bagi perusahaan, dampak finansial dari downtime akibat DDoS sangatlah signifikan, disertai risiko hilangnya kepercayaan pelanggan dan rusaknya reputasi bisnis. Studi kasus menunjukkan kerugian mencapai US$40.000 (sekitar Rp560 juta) per jam dapat diakibatkan oleh serangan DDoS yang melumpuhkan layanan perusahaan [4]. Dengan meningkatnya skala dan kompleksitas serangan DDoS dewasa ini, korban sering kali kesulitan untuk bertahan dari serangan tersebut [3]. Hal ini diperburuk oleh fakta bahwa penyerang terus mengembangkan teknik baru yang mampu menembus atau membanjiri pertahanan tradisional. Oleh karena itu, infrastruktur jaringan enterprise membutuhkan sistem deteksi dan mitigasi DDoS yang canggih dan proaktif untuk mengurangi risiko serangan semacam ini.

Untuk menjawab kebutuhan tersebut, penelitian ini menggunakan dua algoritma machine learning yang telah terbukti efektif dalam tugas klasifikasi, yaitu Random Forest dan CatBoost. Random Forest terkenal dengan kemampuannya menangani data berdimensi tinggi serta menghindari overfitting, sementara CatBoost unggul dalam memproses data kategorikal dan meminimalisir kebutuhan pra-pemrosesan data. Penggunaan kedua algoritma ini secara komparatif bertujuan untuk menemukan pendekatan terbaik dalam mendeteksi serangan DDoS secara akurat dan efisien.

Kondisi nyata di lapangan memperlihatkan urgensi peningkatan keamanan terhadap DDoS. Sejumlah insiden terkini di Indonesia menegaskan seriusnya dampak serangan DDoS terhadap layanan publik maupun korporasi. Komisi Pemilihan Umum (KPU) mengungkapkan bahwa situs resminya mengalami serangan DDoS dalam jumlah ratusan juta kali pada hari pemungutan suara Pemilu 2024, yang menyebabkan gangguan akses terhadap informasi di situs tersebut [5]. Insiden ini terjadi pada momen genting penyelenggaraan pemilu, sehingga menarik perhatian luas dan menunjukkan bahwa bahkan sistem pemerintah kritikal rentan dilumpuhkan oleh DDoS. Kasus lain, pada tahun 2016 situs e-commerce Tiket.com serta website maskapai Citilink diserang DDoS hingga tidak dapat diakses oleh pengguna [3]. Demikian pula, situs web lembaga pemerintah seperti DPR RI pernah tumbang akibat serangan DDoS di tahun 2020, yang memicu kekhawatiran tentang kekuatan pertahanan siber instansi pemerintah [3]. Topologi serangan

DDoS yang menargetkan sistem seperti KPU, Citilink, dan Tiket.com umumnya melibatkan jaringan botnet, yaitu sekumpulan perangkat yang telah dikompromi oleh penyerang. Botnet ini akan mengirimkan trafik palsu secara simultan dalam jumlah besar ke server korban hingga sumber daya sistem tidak mampu merespons permintaan sah. Tujuan utama dari serangan ini adalah untuk menciptakan kondisi *denial-of-service*, yaitu membuat layanan tidak tersedia bagi pengguna sah, baik karena alasan sabotase, gangguan layanan publik, maupun demonstrasi digital oleh pihak-pihak yang tidak bertanggung jawab. Contoh-contoh kasus tersebut menunjukkan bahwa serangan DDoS nyata terjadi di Indonesia dan cenderung meningkat, sehingga kebutuhan akan solusi deteksi dini yang andal menjadi sangat mendesak.

Untuk menghadapi ancaman DDoS yang kian kompleks, pendekatan deteksi serangan berbasis machine learning muncul sebagai salah satu solusi menjanjikan. Berbeda dengan sistem keamanan tradisional (misalnya firewall statis atau sistem deteksi intrusi berbasis tanda tangan) yang sering kali kewalahan menghadapi pola serangan baru, teknik machine learning dapat mempelajari pola lalu lintas normal dan malicious dari data historis dan mendeteksi anomali serangan secara otomatis. Agar deteksi berbasis machine learning efektif, diperlukan dataset yang representatif sebagai bahan pelatihan model. Dalam konteks penelitian ini, digunakan dataset CIC-DDoS2019 (Canadian Institute for Cybersecurity – DDoS 2019) yang merupakan dataset publik berisi trafik jaringan normal dan berbagai jenis serangan DDoS terkini. Dataset CIC-DDoS2019 dipilih karena sifatnya yang komprehensif dan up-to-date; dataset ini mencakup beragam skenario serangan (termasuk beberapa tipe serangan reflektif seperti DNS amplification, UDP flood, HTTP-based DDoS, dan lain-lain) serta trafik benign, sehingga menyerupai pola lalu lintas di dunia nyata [6]. Dengan total lebih dari 12 juta rekaman lalu lintas dan sekitar 80 fitur yang diekstraksi, dataset tersebut berukuran besar dan balanced (sekitar 50% data serangan dan 50% normal) [6]. Hal ini memungkinkan evaluasi model deteksi secara menyeluruh dan reliabel. Melalui dataset inilah algoritma deteksi akan dilatih dan diuji, sehingga diharapkan mampu mengidentifikasi serangan DDoS dengan tingkat akurasi tinggi pada lingkungan jaringan enterprise.

Sejumlah penelitian terdahulu telah mengkaji metode deteksi DDoS menggunakan teknik machine learning pada dataset seperti CIC-DDoS2019. Hasilnya cukup beragam, menunjukkan bahwa tidak ada satu pendekatan tunggal yang selalu superior, dan hal ini membuka peluang eksplorasi lebih lanjut. Misalnya, sebuah studi menggunakan algoritma Extreme Learning Machine yang

dioptimasi dengan algoritme blackhole mampu mencapai akurasi deteksi hingga 99,80% [6]. Di sisi lain, pendekatan berbasis deep neural network (DNN) menghasilkan akurasi sekitar 94,57% [6] pada dataset yang sama. Algoritma klasik seperti decision tree dilaporkan mencapai akurasi 92% saja [6], kalah dibandingkan metode yang lebih canggih. Menariknya, dalam suatu studi komparatif yang melibatkan beberapa algoritma klasifikasi (Artificial Neural Network, SVM, K-Nearest Neighbor, dll.), algoritma sederhana Naïve Bayes justru tercatat memberikan hasil prediksi terbaik [6]. Sementara itu, metode Random Forest – yang merupakan ensembel banyak pohon keputusan – sering diakui handal untuk deteksi intrusi; sebuah penelitian bahkan mencatat model Random Forest mampu meraih akurasi 99,997% pada CIC-DDoS2019 setelah dilakukan optimasi fitur dan parameter [6]. Variasi temuan ini menunjukkan bahwa berbagai pendekatan machine learning telah dicoba untuk mendeteksi DDoS, namun masing-masing punya kelebihan dan kekurangan. Banyak penelitian cenderung hanya menggunakan satu algoritma tertentu atau membandingkan model-model yang berbeda jenis (misalnya pembandingan antara pohon keputusan dan jaringan saraf). Belum banyak riset yang secara spesifik membandingkan dua algoritma ensemble berbasis pohon keputusan dalam skenario deteksi DDoS yang sama. Dengan kata lain, terdapat knowledge gap di mana efektivitas relatif antara algoritma ensemble bagging seperti Random Forest dan algoritma ensemble boosting modern seperti CatBoost belum diteliti secara mendalam pada domain deteksi DDoS.

Berdasarkan paparan di atas, penelitian ini diusulkan untuk mengisi kesenjangan pengetahuan tersebut dengan melakukan implementasi dan evaluasi dua algoritma machine learning berbasis ensemble decision tree, yaitu Random Forest dan CatBoost, dalam mendeteksi serangan DDoS. CatBoost merupakan algoritma gradient boosting berbasis pohon keputusan yang tergolong baru dan dikenal mampu mencapai akurasi tinggi di berbagai tugas klasifikasi [6], namun penggunaannya dalam deteksi DDoS masih terbatas dalam literatur. Sementara itu, Random Forest adalah algoritma ensemble berbasis bagging yang sudah mapan dan kerap menunjukkan kinerja unggul dalam deteksi anomali jaringan. Dengan menerapkan kedua algoritma ini pada dataset CIC-DDoS2019, penelitian ini bertujuan untuk mengevaluasi kinerja model secara komprehensif berdasarkan metrik precision, recall, dan F1-score. Ketiga metrik tersebut dipilih karena mampu menggambarkan keseimbangan kemampuan deteksi (baik dalam mengidentifikasi serangan maupun menghindari alarm salah) secara lebih detail dibanding sekadar akurasi. Melalui evaluasi ini, akan dibandingkan keunggulan masing-masing

algoritma dalam mendeteksi berbagai pola serangan DDoS.

Target yang ingin dicapai adalah diperolehnya model deteksi DDoS yang optimal untuk infrastruktur jaringan enterprise, atau setidaknya pemahaman yang lebih jelas tentang algoritma mana (Random Forest atau CatBoost) yang lebih sesuai untuk tugas tersebut. Adapun kontribusi ilmiah dari penelitian ini adalah memberikan wawasan baru mengenai perbandingan dua algoritma ensemble learning terkini dalam domain keamanan jaringan. Hasil penelitian diharapkan dapat memperkuat kesinambungan studi terdahulu dengan menambahkan analisis perbandingan yang belum banyak dilakukan, sekaligus menghadirkan novelty berupa rekomendasi algoritma atau konfigurasi model yang dapat meningkatkan akurasi deteksi serangan DDoS. Dengan demikian, penelitian ini tidak hanya relevan secara akademis, tetapi juga bernilai praktis dalam upaya meningkatkan ketahanan infrastruktur jaringan enterprise di Indonesia terhadap ancaman serangan DDoS yang kian meningkat.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini disusun berdasarkan latar belakang yang telah diuraikan, sebagaimana ditunjukkan dalam poin-poin berikut:

1. Bagaimana kinerja algoritma Random Forest dalam mendeteksi serangan DDoS menggunakan dataset CIC-DDoS2019?
2. Bagaimana kinerja algoritma CatBoost dalam mendeteksi serangan DDoS menggunakan dataset CIC-DDoS2019?
3. Bagaimana perbandingan kinerja algoritma Random Forest dan CatBoost dalam mendeteksi serangan DDoS jika diukur berdasarkan metrik precision, recall, dan F1-score?
4. Algoritma manakah yang memberikan hasil deteksi paling optimal dan stabil dalam skenario deteksi serangan DDoS?

1.3 Batasan Permasalahan

Adapun batasan permasalahan dari penelitian ini adalah:

1. **Jenis Serangan**
   Penelitian ini hanya memfokuskan pada deteksi serangan Distributed Denial

of Service (DDoS) yang terdapat dalam dataset CIC-DDoS2019, tanpa membahas jenis serangan siber lainnya seperti malware, phishing, atau SQL injection.

2. **Dataset yang Digunakan**
   Data yang digunakan dalam penelitian ini adalah dataset CIC-DDoS2019 yang diperoleh dari Canadian Institute for Cybersecurity. Dataset ini sudah ditentukan sebelumnya dan tidak dilakukan pengumpulan data langsung dari jaringan enterprise secara real-time.

3. **Metode Deteksi**
   Penelitian ini membandingkan dua algoritma machine learning, yaitu:

   - Random Forest (ensemble berbasis bagging), dan
   - CatBoost (ensemble berbasis boosting).

   Algoritma lain seperti SVM, KNN, atau neural network tidak dibahas dalam penelitian ini.

4. **Evaluasi Kinerja**
   Evaluasi kinerja deteksi model hanya dilakukan dengan menggunakan tiga metrik utama, yaitu:

   - Precision
   - Recall
   - F1-Score

   Pengukuran akurasi, ROC-AUC, atau metrik lainnya tidak dibahas secara mendalam.

5. **Skala dan Lingkungan Pengujian**
   Penelitian ini bersifat eksperimental dan seluruh proses dilakukan dalam lingkungan simulasi menggunakan dataset publik. Tidak dilakukan implementasi langsung pada sistem jaringan enterprise nyata.

6. **Pra-pemrosesan Data**
   Fokus penelitian adalah pada tahap pelatihan dan evaluasi model, sehingga tahap pra-pemrosesan data (seperti normalisasi, seleksi fitur, dan balancing) hanya dilakukan seperlunya untuk menyiapkan data sebelum pelatihan model, tanpa menjadi fokus utama penelitian.

7. **Variabel Bebas dan Terikat**

   - Variabel bebas dalam penelitian ini adalah jenis algoritma yang digunakan (Random Forest dan CatBoost).
   - Variabel terikat adalah nilai precision, recall, dan F1-score yang dihasilkan dari masing-masing algoritma.

1.4 Tujuan Penelitian

Adapun tujuan dilakukannya penelitian pada skripsi ini adalah:

1. Mengetahui kinerja algoritma Random Forest dalam mendeteksi serangan DDoS menggunakan dataset CIC-DDoS2019 pada infrastruktur jaringan enterprise.
2. Mengetahui kinerja algoritma CatBoost dalam mendeteksi serangan DDoS menggunakan dataset CIC-DDoS2019 pada infrastruktur jaringan enterprise.
3. Melakukan perbandingan kinerja antara algoritma Random Forest dan CatBoost berdasarkan metrik precision, recall, dan F1-score.
4. Menentukan algoritma yang paling optimal dalam hal akurasi deteksi serangan DDoS pada lingkungan jaringan yang disimulasikan.

1.5 Manfaat Penelitian

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

1. Memberikan kontribusi ilmiah berupa studi komparatif antara algoritma Random Forest dan CatBoost dalam konteks deteksi serangan DDoS berbasis machine learning.
2. Menambah wawasan dan literatur di bidang keamanan jaringan dan pembelajaran mesin, khususnya terkait penerapan algoritma ensemble untuk pendeteksian ancaman siber.
3. Menyediakan referensi praktis bagi praktisi dan profesional TI dalam memilih model deteksi DDoS yang lebih efektif untuk diterapkan pada sistem keamanan jaringan enterprise.

4. Membantu dalam pengembangan sistem keamanan jaringan yang lebih adaptif dan efisien, khususnya dalam menghadapi serangan DDoS yang kompleks dan berkembang.

1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
  Bab ini memuat latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan skripsi. Bagian ini bertujuan memberikan gambaran awal mengenai urgensi dan fokus dari penelitian yang dilakukan.

- Bab 2 LANDASAN TEORI
  Bab ini membahas teori-teori yang relevan dengan penelitian, termasuk pengertian serangan DDoS, penjelasan dataset CIC-DDoS2019, prinsip kerja algoritma Random Forest dan CatBoost, serta metode evaluasi seperti precision, recall, dan F1-score. Landasan ini menjadi dasar dalam proses analisis dan implementasi model.

- Bab 3 METODOLOGI PENELITIAN
  Bab ini menjelaskan langkah-langkah penelitian secara sistematis, mulai dari pengumpulan data, preprocessing, pemilihan fitur, pembagian data, hingga tahapan pelatihan dan pengujian model dengan algoritma yang digunakan.

- Bab 4 HASIL DAN DISKUSI
  Bab ini menyajikan hasil evaluasi dari model yang telah dilatih, disertai dengan analisis performa kedua algoritma berdasarkan metrik evaluasi. Hasilnya dibandingkan untuk memperoleh kesimpulan mengenai algoritma yang lebih optimal dalam mendeteksi serangan DDoS.

- Bab 5 SIMPULAN DAN SARAN
  Bab terakhir ini berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan atau penelitian lanjutan di masa depan.