PERBANDINGAN KINERJA ALGORITMA RANDOM FOREST DAN CATBOOST DALAM DETEKSI SERANGAN DDOS MENGGUNAKAN DATASET CIC-DDOS2019

SKRIPSI

EZEKIEL CHRISTOFORUS ALLEGRA HANVERLOY SHINDORO
00000057001

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

PERBANDINGAN KINERJA ALGORITMA RANDOM FOREST DAN CATBOOST DALAM DETEKSI SERANGAN DDOS MENGGUNAKAN DATASET CIC-DDOS2019

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

EZEKIEL CHRISTOFORUS ALLEGRA HANVERLOY SHINDORO
00000057001

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Ezekiel Christoforus Allegra Hanverloy Shindoro
Nomor Induk Mahasiswa	:	00000057001
Program Studi	:	Informatika

Skripsi dengan judul:

Perbandingan Kinerja Algoritma Random Forest dan CatBoost dalam Deteksi Serangan DDoS Menggunakan Dataset CIC-DDoS2019

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 26 Juni 2025

(Ezekiel Christoforus Allegra Hanverloy Shindoro)

HALAMAN PENGESAHAN

Skripsi dengan judul

PERBANDINGAN KINERJA ALGORITMA RANDOM FOREST DAN CATBOOST DALAM DETEKSI SERANGAN DDOS MENGGUNAKAN DATASET CIC-DDOS2019

oleh

Nama	:	Ezekiel Christoforus Allegra Hanverloy Shindoro
NIM	:	00000057001
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Senin, 21 Juli 2025
Pukul 10.00 s/s 12.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(David Agustriawan, S.Kom., M.Sc., Ph.D.)
NIDN: 0525088601

Penguji

(Moeljono Widjaja, B.Sc., M.Sc., Ph.D)
NIDN: 0311106903

Pembimbing

(Angga Aditya Permana, S.Kom., M.Kom.)
NIDN: 0407128901

Ketua Program Studi Informatika,

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)
NIDN: 0315109103

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	:	Ezekiel Christoforus Allegra Hanverloy Shindoro
NIM	:	00000057001
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Perbandingan Kinerja Algoritma Random Forest dan CatBoost dalam Deteksi Serangan DDoS Menggunakan Dataset CIC-DDoS2019

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

☑ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.

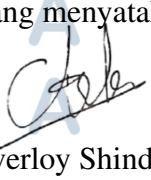
☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.

☐ Lainnya, pilih salah satu:

- Hanya dapat diakses secara internal Universitas Multimedia Nusantara
- Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, 26 Juni 2025

Yang menyatakan



Ezekiel Christoforus Allegra Hanverloy Shindoro

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

HALAMAN PERSEMBAHAN / MOTTO

”I can do all things through him who strengthens me.”

Philippians 4:13

KATA PENGANTAR

Puji dan syukur ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga karya ilmiah ini dapat diselesaikan. Skripsi ini disusun untuk menganalisis dan membandingkan performa dua algoritma pembelajaran mesin, yaitu Random Forest dan CatBoost, dalam mendeteksi serangan Distributed Denial of Service (DDoS) menggunakan dataset CIC-DDoS2019.

Tanpa bimbingan, bantuan, dan dukungan dari berbagai pihak, penyusunan skripsi ini tidak akan berjalan dengan baik. Oleh karena itu, diucapkan terima kasih kepada:

1. Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Angga Aditya Permana, S.Kom., M.Kom., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga yang telah memberikan bantuan dukungan material dan moral, sehingga dapat menyelesaikan tugas akhir ini.

Skripsi ini tentu masih memiliki kekurangan. Oleh karena itu, saran dan kritik yang membangun sangat diharapkan demi penyempurnaan karya ini. Akhir kata, semoga karya ilmiah ini dapat memberikan manfaat bagi dunia akademik, serta pengembangan teknologi keamanan jaringan di masa mendatang.

Tangerang, 26 Juni 2025

Ezekiel Christoforus Allegra Hanverloy Shindoro

PERBANDINGAN KINERJA ALGORITMA RANDOM FOREST DAN CATBOOST DALAM DETEKSI SERANGAN DDOS MENGGUNAKAN DATASET CIC-DDOS2019

Ezekiel Christoforus Allegra Hanverloy Shindoro

ABSTRAK

Serangan *Distributed Denial of Service* (DDoS) merupakan ancaman serius bagi infrastruktur jaringan enterprise. Penelitian ini membandingkan kinerja algoritma Random Forest dan CatBoost dalam mendeteksi serangan DDoS menggunakan dataset CIC-DDoS2019. Dataset yang digunakan terdiri dari enam kelas, yaitu BENIGN, LDAP, MSSQL, NetBIOS, Portmap, dan SYN, masing-masing berjumlah 200 sampel. Evaluasi dilakukan berdasarkan metrik *precision*, *recall*, *f1-score*, dan waktu pelatihan. Hasil penelitian menunjukkan bahwa Random Forest unggul dalam efisiensi pelatihan dan stabilitas klasifikasi, terutama pada label BENIGN dan SYN. Di sisi lain, CatBoost memperoleh skor rata-rata lebih tinggi melalui validasi silang 5-fold, yang mengindikasikan kemampuan generalisasi yang lebih baik. Dengan demikian, CatBoost lebih cocok untuk deteksi multikelas yang akurat, sedangkan Random Forest lebih efisien digunakan dalam kondisi yang memprioritaskan kecepatan komputasi.

Kata kunci: CatBoost, DDoS, Deteksi Intrusi, Random Forest, Trafik Jaringan

COMPARATIVE PERFORMANCE OF RANDOM FOREST AND CATBOOST ALGORITHMS IN DETECTING DDOS ATTACKS USING THE CIC-DDOS2019 DATASET

Ezekiel Christoforus Allegra Hanverloy Shindoro

ABSTRACT

Distributed Denial of Service (DDoS) attacks pose a serious threat to enterprise network infrastructures. This study compares the performance of Random Forest and CatBoost algorithms in detecting DDoS attacks using the CIC-DDoS2019 dataset. The dataset consists of six classes: BENIGN, LDAP, MSSQL, NetBIOS, Portmap, and SYN, each containing 200 samples. Evaluation is based on precision, recall, f1-score, and training time. The results show that Random Forest outperforms in training efficiency and classification stability, especially on the BENIGN and SYN labels. On the other hand, CatBoost achieves higher average scores in 5-fold cross-validation, indicating better generalization capabilities. Therefore, CatBoost is more suitable for accurate multi-class detection, while Random Forest is preferable in scenarios prioritizing computational speed.

Keywords: *CatBoost, DDoS, Intrusion Detection, Network Traffic, Random Forest*

DAFTAR ISI

HALAMAN JUDUL i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT ii
HALAMAN PENGESAHAN iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH iv
HALAMAN PERSEMBAHAN/MOTO v
KATA PENGANTAR vi
ABSTRAK vii
ABSTRACT viii
DAFTAR ISI ix
DAFTAR TABEL xi
DAFTAR GAMBAR xii
DAFTAR KODE xiii
DAFTAR RUMUS xiv
DAFTAR LAMPIRAN xv
BAB 1 PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 5
1.3 Batasan Permasalahan 5
1.4 Tujuan Penelitian 7
1.5 Manfaat Penelitian 7
1.6 Sistematika Penulisan 8
BAB 2 LANDASAN TEORI 9
2.1 Serangan Distributed Denial of Service (DDoS) 9
2.2 Deteksi Serangan Berbasis Machine Learning 10
2.3 Random Forest 10
2.4 CatBoost 12
2.5 Metode Evaluasi Kinerja 14
2.5.1 Precision 14
2.5.2 Recall 14
2.5.3 F1-Score 15
2.6 Feature Extraction 15
2.7 Feature Importance 16
2.8 5-Fold Cross-Validation 17
2.9 Dataset CIC-DDoS2019 17
BAB 3 METODOLOGI PENELITIAN 19
3.1 Jenis dan Pendekatan Penelitian 19
3.2 Alur Penelitian 19
3.3 Tools dan Lingkungan Pengujian 21
BAB 4 HASIL DAN DISKUSI 22
4.1 Deskripsi Dataset 22
4.2 Preprocessing Data 23
4.2.1 Penggabungan File Dataset 24
4.2.2 Pembersihan Kolom Tidak Relevan 25
4.2.3 Penanganan Nilai Hilang dan Nilai Tak Hingga 25
4.2.4 Seleksi Fitur Penting 25
4.2.5 Sampling Data dan Encoding Label 28
4.2.6 Pembagian Data Latih dan Uji 29
4.3 Implementasi dan Evaluasi Model 29

4.3.1 Random Forest 29
4.3.2 CatBoost 30
4.4 Perbandingan Kinerja Model 31
4.5 Visualisasi Heatmap Perbandingan Skor Per Kelas 32
4.6 Analisis Confusion Matrix 33
4.7 Analisis Feature Importance 34
4.8 Evaluasi Model dengan 5-Fold Cross-Validation 36
4.9 Analisis Mendalam Hasil Eksperimen 39
BAB 5 SIMPULAN DAN SARAN 41
5.1 Kesimpulan 41
5.2 Saran 42
DAFTAR PUSTAKA 43

DAFTAR TABEL

Tabel 4.1	Rincian frekuensi data setelah proses sampling	22
Tabel 4.2	Penjelasan label serangan pada dataset	23
Tabel 4.3	Penjelasan fitur penting yang digunakan	26
Tabel 4.4	Perbandingan f1-score per label antara Random Forest dan CatBoost	31
Tabel 4.5	F1-score pada setiap fold untuk Random Forest dan CatBoost	37
Tabel 4.6	Precision dan recall setiap fold pada Random Forest dan CatBoost	38

DAFTAR GAMBAR

Gambar 1.1 Distribusi serangan DDoS secara global dan posisi Indonesia 1
Gambar 2.1 Diagram alur DDoS 9
Gambar 2.2 Alur algoritma Random Forest 11
Gambar 2.3 Alur algoritma CatBoost 13
Gambar 3.1 Diagram alur penelitian 19
Gambar 4.1 Heatmap perbandingan precision, recall, dan f1-score antara Random Forest dan CatBoost untuk setiap kelas . . 32
Gambar 4.2 Confusion matrix untuk algoritma Random Forest 33
Gambar 4.3 Confusion matrix untuk algoritma CatBoost 34
Gambar 4.4 Feature importance pada Random Forest 35
Gambar 4.5 Feature importance pada CatBoost 35
Gambar 4.6 Perbandingan F1-score per fold antara Random Forest dan CatBoost 37
Gambar 4.7 Visualisasi line chart precision dan recall CatBoost dan Random Forest 38

DAFTAR KODE

Kode 4.1 Penggabungan dan pemisahan label serangan 24
Kode 4.2 Balancing jumlah sampel per label 24
Kode 4.3 Menghapus kolom non-informatif 25
Kode 4.4 Menangani nilai NaN dan Inf 25
Kode 4.5 Untuk seleksi fitur 28
Kode 4.6 Sampling data dan encoding label 28
Kode 4.7 Pembagian data latih dan uji 29
Kode 4.8 Pelatihan dan evaluasi model Random Forest 30
Kode 4.9 Pelatihan dan evaluasi CatBoost 30

DAFTAR RUMUS

Rumus 2.1 *Random Forest* 11
Rumus 2.2 *CatBoost* 13
Rumus 2.3 *Precision* 14
Rumus 2.4 *Recall* 15
Rumus 2.5 *F1 Score* 15

DAFTAR LAMPIRAN

Lampiran 1 Turnitin 45
Lampiran 2 Formulir Bimbingan 52