Published online on the page: https://journal.upp.ac.id/index.php/rjti/index

R J T I
Riau Jurnal Teknik Informatika
| E - ISSN (*Online*) 2829-2529 |

Deteksi Serangan *DDoS* Menggunakan *Explainable Ensemble Learning* dan Analisis SHAP

Sutrisno[1,*], Okta Irawati[2]

[1,2] Universitas Pamulang, Tanggerang Selatan, Indonesia

Informasi Artikel

Sejarah Artikel:
Submit: 21 Juni 2026
Revisi: 05 Juli 2026
Diterima: 14 Juli 2026
Diterbitkan: 19 Juli 2026

Kata Kunci

Explainable AI, DDoS, SHAP, Ensemble Learning, Intrusion Detection System.

Korespondensi

E-mail: dosen02673@unpa

A B S T R A C T

Distributed Denial of Service (DDoS) attacks remain a major threat to network service availability due to their ability to generate massive traffic volumes that closely resemble legitimate activities. This study proposes an Explainable Ensemble Learning approach for DDoS detection using the CIC-DDoS2019 dataset. The proposed framework integrates Mutual Information-based feature selection to identify the 20 most relevant features, Synthetic Minority Over-sampling Technique (SMOTE) for class balancing, and a Voting Ensemble of Random Forest, XGBoost, and LightGBM classifiers. Model performance was evaluated using a 70:30 train-test split with Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics. Experimental results achieved 99.80% Accuracy, 99.79% F1-score, and 0.9999 ROC-AUC. SHAP analysis identified Avg_Packet_Size and packet-length-related features as the most influential predictors, improving both detection performance and model interpretability. These findings demonstrate that integrating Ensemble Learning with Explainable Artificial Intelligence provides an accurate and transparent solution for DDoS detection interpretability.

Abstrak

Serangan Distributed Denial of Service (DDoS) masih menjadi salah satu ancaman utama terhadap ketersediaan layanan jaringan karena mampu menghasilkan lalu lintas dalam volume besar yang sulit dibedakan dari aktivitas normal. Penelitian ini mengusulkan pendekatan Explainable Ensemble Learning untuk mendeteksi serangan DDoS menggunakan dataset CIC-DDoS2019. Kerangka kerja yang dikembangkan mengintegrasikan seleksi fitur berbasis Mutual Information untuk memilih 20 fitur paling relevan, penyeimbangan data menggunakan Synthetic Minority Over-sampling Technique (SMOTE), serta Voting Ensemble yang terdiri atas Random Forest, XGBoost, dan LightGBM. Evaluasi model dilakukan menggunakan train-test split 70:30 dengan metrik Accuracy, Precision, Recall, F1-score, dan ROC-AUC. Hasil eksperimen menunjukkan Accuracy sebesar 99,80%, F1-score 99,79%, dan ROC-AUC 0,9999. Analisis SHAP mengidentifikasi Avg_Packet_Size dan fitur-fitur berbasis panjang paket sebagai prediktor paling berpengaruh, sehingga meningkatkan performa deteksi sekaligus interpretabilitas model. Hasil penelitian menunjukkan bahwa integrasi Ensemble Learning dan Explainable Artificial Intelligence mampu menghasilkan sistem deteksi DDoS yang akurat, andal, dan transparan.

1. Pendahuluan

Transformasi digital yang berlangsung pada berbagai sektor telah meningkatkan ketergantungan organisasi terhadap layanan jaringan dan komputasi berbasis internet. Ketersediaan layanan menjadi faktor kritis karena gangguan dalam waktu singkat dapat menyebabkan kerugian operasional, finansial, maupun reputasi. Dalam konteks tersebut, serangan Distributed Denial of Service (DDoS) masih menjadi salah satu ancaman siber yang paling sering ditemukan dan terus berkembang baik dari sisi skala, kompleksitas, maupun teknik yang digunakan [1], [2]. Pemanfaatan botnet, perangkat Internet of Things (IoT), dan infrastruktur cloud memungkinkan penyerang menghasilkan lalu lintas dalam jumlah sangat besar sehingga mampu melumpuhkan layanan jaringan dalam waktu singkat [3].

Karakteristik serangan DDoS modern berbeda dibandingkan dengan generasi sebelumnya. Serangan tidak lagi hanya berfokus pada peningkatan volume trafik, tetapi juga memanfaatkan teknik refleksi, amplifikasi, dan penyamaran pola lalu lintas agar menyerupai aktivitas pengguna yang sah Akibatnya, mekanisme keamanan tradisional berbasis signature dan rule-based mengalami keterbatasan dalam mengenali pola serangan baru yang belum pernah terdaftar sebelumnya [4]. Tantangan ini semakin kompleks ketika lingkungan jaringan harus menangani lalu lintas dengan volume tinggi dan pola komunikasi yang dinamis.

Perkembangan Machine Learning memberikan peluang baru dalam membangun sistem deteksi intrusi yang lebih adaptif. Berbeda dengan pendekatan konvensional, algoritma Machine Learning mampu mempelajari karakteristik lalu lintas jaringan secara otomatis dan mengidentifikasi pola anomali berdasarkan data historis [5]. Berbagai penelitian melaporkan keberhasilan penggunaan Random Forest, Support Vector Machine, XGBoost, hingga Deep Learning untuk meningkatkan akurasi deteksi serangan DDoS [6]. Meskipun demikian, peningkatan performa klasifikasi belum sepenuhnya menyelesaikan permasalahan yang dihadapi pada implementasi nyata.

Sebagian besar model dengan akurasi tinggi masih bekerja sebagai *black box*, yaitu menghasilkan prediksi tanpa memberikan penjelasan yang memadai mengenai alasan di balik keputusan tersebut [7]. Dalam lingkungan operasional keamanan siber, transparansi menjadi aspek penting karena administrator jaringan perlu memahami faktor-faktor yang menyebabkan suatu koneksi dikategorikan sebagai serangan. Kurangnya interpretabilitas dapat mengurangi tingkat kepercayaan terhadap sistem deteksi dan menyulitkan proses investigasi insiden [8].

Di sisi lain, penggunaan model tunggal juga memiliki keterbatasan dalam menghadapi variasi karakteristik serangan. Random Forest dikenal stabil terhadap data berdimensi tinggi, XGBoost memiliki kemampuan generalisasi yang kuat, sedangkan LightGBM menawarkan efisiensi komputasi yang lebih baik pada dataset berskala besar [9]. Masing-masing algoritma memiliki keunggulan dan kelemahan yang berbeda, sehingga tidak selalu memberikan performa optimal pada seluruh kondisi lalu lintas jaringan. Pendekatan Ensemble Learning menjadi alternatif yang menarik karena mampu menggabungkan kekuatan beberapa model sekaligus untuk menghasilkan keputusan yang lebih robust [10].

Sejalan dengan perkembangan tersebut, Explainable Artificial Intelligence (XAI) mulai banyak digunakan untuk meningkatkan transparansi model Machine Learning. Salah satu metode yang paling banyak diterapkan adalah SHapley Additive exPlanations (SHAP), yang mampu menjelaskan kontribusi setiap fitur terhadap keputusan model berdasarkan konsep nilai Shapley dalam teori permainan kooperatif [11]. Informasi ini sangat penting pada domain keamanan siber karena membantu analis memahami indikator lalu lintas yang paling berpengaruh terhadap deteksi serangan.

Penelitian ini menggunakan dataset CIC-DDoS2019 yang dikembangkan oleh Canadian Institute for Cybersecurity sebagai salah satu benchmark paling komprehensif untuk penelitian DDoS [12].

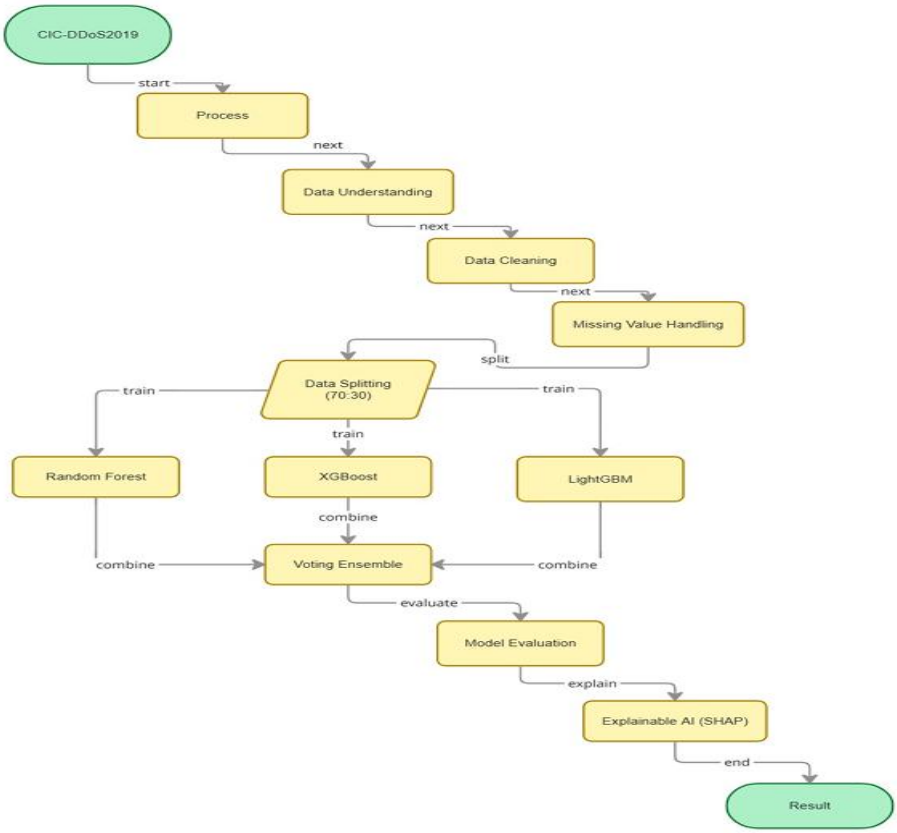
Dataset tersebut mencakup berbagai jenis serangan modern, termasuk DrDoS_NTP, DNS, LDAP, MSSQL, NetBIOS, SNMP, SSDP, SYN, TFTP, dan UDP Flood. Analisis awal menunjukkan adanya ketidakseimbangan distribusi kelas serta korelasi tinggi pada beberapa fitur lalu lintas jaringan. Kondisi tersebut memerlukan strategi preprocessing yang tepat agar model mampu mempelajari pola data secara optimal.

Kajian terhadap penelitian terdahulu menunjukkan bahwa sebagian besar studi masih berfokus pada peningkatan akurasi klasifikasi tanpa mengintegrasikan aspek interpretabilitas secara mendalam [13] Penelitian yang menggabungkan Explainable AI dengan Ensemble Learning untuk deteksi DDoS juga masih relatif terbatas, terutama pada dataset CIC-DDoS2019 [14]. Selain itu, belum banyak penelitian yang mengombinasikan seleksi fitur berbasis Mutual Information, penyeimbangan data menggunakan SMOTE, serta interpretasi model menggunakan SHAP dalam satu kerangka kerja terpadu.

2. Metode Penelitian

2.1. Tahapan Penelitian

Penelitian ini mengusulkan pendekatan **Explainable Ensemble Learning** untuk mendeteksi serangan Distributed Denial of Service (DDoS) menggunakan dataset CIC-DDoS2019. Kerangka penelitian mencakup tahapan preprocessing data, seleksi fitur, penanganan ketidakseimbangan kelas, pembangunan model klasifikasi, evaluasi performa, dan interpretasi hasil menggunakan Explainable Artificial Intelligence (XAI).



Gambar 1. Tahapan Penelitian Explainable Ensemble Learning

Dataset CIC-DDoS2019 dipilih karena menyediakan representasi lalu lintas jaringan yang beragam, mencakup trafik normal dan berbagai jenis serangan DDoS modern seperti DrDoS_NTP, DNS, LDAP, MSSQL, NetBIOS, SNMP, SSDP, SYN, TFTP, dan UDP Flood [2]. Karakteristik tersebut menjadikan dataset ini relevan untuk mengevaluasi kemampuan model dalam menghadapi kondisi

jaringan yang kompleks. Tahapan penelitian secara umum ditunjukkan pada Gambar 1. Alur tersebut menggambarkan hubungan antarproses mulai dari pengolahan data hingga interpretasi hasil klasifikasi.

Tahapan penelitian ditunjukkan pada Gambar 1, yang menggambarkan alur mulai dari pengolahan data hingga interpretasi hasil klasifikasi.

2.2. Preprocessing dan Seleksi Fitur

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum proses pembelajaran mesin. Proses ini meliputi penanganan *missing values*, penghapusan data duplikat, transformasi label, serta normalisasi fitur menggunakan StandardScaler [13]. Normalisasi diperlukan karena fitur pada CIC-DDoS2019 memiliki rentang nilai yang berbeda sehingga berpotensi memengaruhi proses pelatihan model [6].

Setelah preprocessing, dilakukan seleksi fitur menggunakan Mutual Information (MI) untuk mengukur tingkat ketergantungan antara fitur dan variabel target [6]. Semakin tinggi nilai MI, semakin besar kontribusi fitur terhadap proses klasifikasi. Nilai Mutual Information dihitung menggunakan Persamaan (1).

$$MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

Keterangan:

M I (X; Y) adalah nilai Mutual Information antara fitur X dan label target Y.
p(x,y) adalah probabilitas gabungan (*joint probability*) antara X dan Y.
p(x) adalah probabilitas marginal dari fitur X.
p(y) adalah probabilitas marginal dari label Y.

Hasil proses ini digunakan untuk memilih 20 fitur terbaik yang selanjutnya menjadi masukan pada tahap klasifikasi. Strategi tersebut tidak hanya mengurangi dimensi data, tetapi juga membantu meningkatkan efisiensi komputasi dan kemampuan generalisasi model [15].

2.3. Penanganan Ketidakseimbangan Data

Distribusi kelas pada dataset menunjukkan adanya ketidakseimbangan jumlah sampel antara trafik normal dan serangan. Kondisi ini dapat menyebabkan model lebih cenderung mengenali kelas mayoritas dibandingkan kelas minoritas [7].

Untuk mengatasi masalah tersebut digunakan Synthetic Minority Over-sampling Technique (SMOTE). Metode ini menghasilkan sampel sintetis berdasarkan kedekatan data pada kelas minoritas sehingga distribusi data menjadi lebih seimbang [8]. Penggunaan SMOTE terbukti efektif dalam meningkatkan sensitivitas model pada berbagai penelitian deteksi intrusi [9].

2.4. Pembangunan Model Explainable Ensemble Learning

Model klasifikasi dibangun menggunakan tiga algoritma Machine Learning, yaitu Random Forest, XGBoost, dan LightGBM. Random Forest dipilih karena stabil terhadap data berdimensi tinggi dan mampu mengurangi overfitting melalui teknik bagging [9]. XGBoost memanfaatkan pendekatan gradient boosting untuk meningkatkan kemampuan generalisasi model [10], sedangkan LightGBM menawarkan efisiensi komputasi yang tinggi melalui teknik histogram-based learning [11].

Ketiga model tersebut digabungkan menggunakan metode **Soft Voting Ensemble** untuk menghasilkan keputusan yang lebih robust. Probabilitas akhir dihitung menggunakan Persamaan (2).

$$P(y_k) = \frac{1}{M} \sum_{i=1}^{M} P_i(y_k)$$

dengan (P(y)) menyatakan probabilitas akhir kelas target, (P_i(y)) merupakan probabilitas yang dihasilkan model ke-(i), dan (n) adalah jumlah model dalam ensemble [13]. Pendekatan ini memungkinkan sistem memanfaatkan keunggulan masing-masing algoritma sekaligus mengurangi risiko kesalahan yang muncul pada model tunggal.

2.5. Evaluasi Model dan Explainable AI

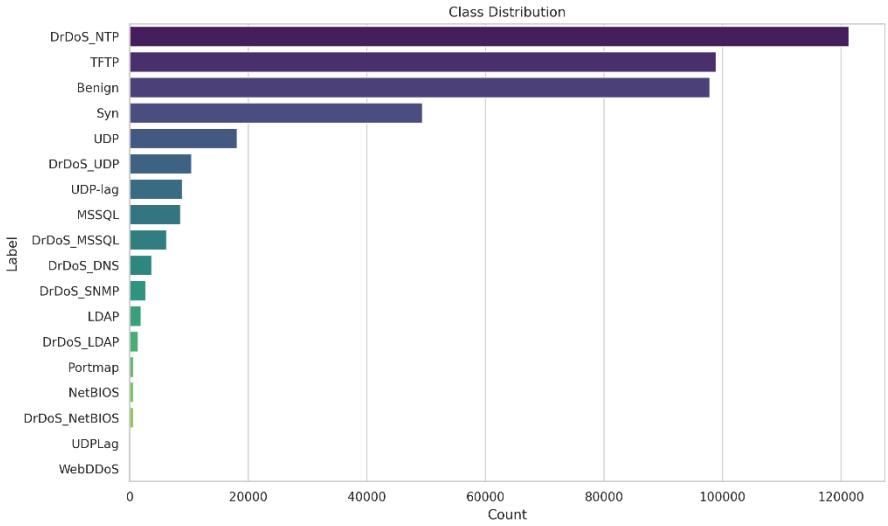
Evaluasi dilakukan menggunakan skema Train-Test Split 70:30 dengan stratifikasi kelas [13]. Kinerja model diukur menggunakan Accuracy, Precision, Recall, F1-Score, ROC-AUC, dan Confusion Matrix [16].

Untuk meningkatkan interpretabilitas, penelitian ini menerapkan SHapley Additive exPlanations (SHAP). Metode SHAP menghitung kontribusi setiap fitur terhadap keputusan model berdasarkan teori permainan kooperatif [17]. Melalui pendekatan ini, fitur-fitur yang paling berpengaruh terhadap deteksi DDoS dapat diidentifikasi secara kuantitatif, sehingga model tidak hanya menghasilkan performa klasifikasi yang tinggi tetapi juga mampu memberikan penjelasan yang transparan terhadap proses pengambilan keputusan [18].

3. Hasil dan Pembahasan

3.1. Analisis Distribusi Data

Tahap awal penelitian difokuskan pada pemahaman karakteristik dataset CIC-DDoS2019 untuk mengetahui tingkat kompleksitas permasalahan yang akan dihadapi model klasifikasi. Distribusi kelas ditunjukkan pada Gambar 2, sedangkan ringkasan jumlah sampel setiap kategori disajikan pada Tabel 1.



Gambar 2. Distribusi Kelas Dataset CIC-DDoS2019

Hasil analisis menunjukkan bahwa distribusi data tidak seimbang. Kelas DrDoS_NTP mendominasi dataset dengan proporsi sekitar 28,1%, diikuti oleh TFTP dan Benign yang masing-masing mencapai sekitar 23%. Sebaliknya, kategori seperti WebDDoS, UDPLag, dan DrDoS_NetBIOS hanya memiliki jumlah sampel yang sangat kecil.

Kondisi tersebut mencerminkan karakteristik lalu lintas jaringan nyata, di mana tidak semua jenis serangan muncul dengan frekuensi yang sama. Serangan berbasis refleksi seperti NTP dan TFTP

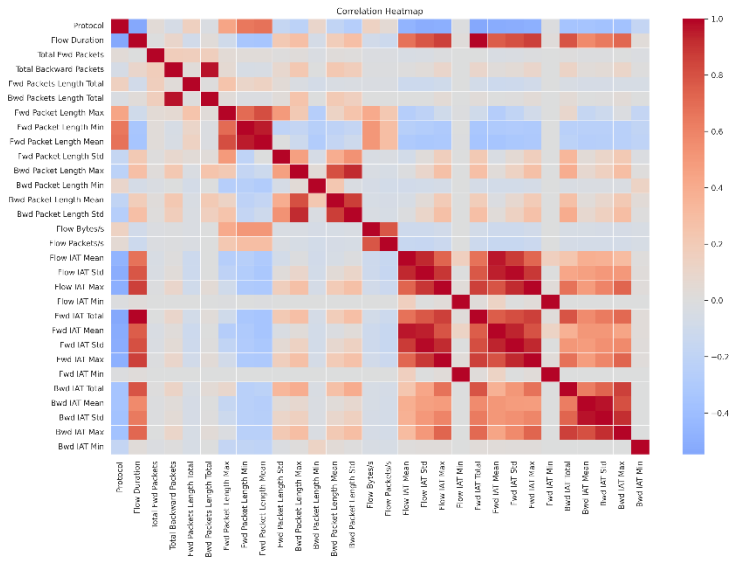
diketahui memiliki kemampuan amplifikasi trafik yang tinggi sehingga lebih sering ditemukan dalam berbagai studi keamanan jaringan. Dari perspektif machine learning, ketidakseimbangan kelas dapat menyebabkan model cenderung mempelajari pola kelas mayoritas dan mengabaikan karakteristik kelas minoritas. Oleh karena itu, penanganan class imbalance menjadi tahapan penting sebelum proses pelatihan dilakukan.

Tabel 1. Distribusi Kelas pada Dataset CIC-DDoS2019

No	Kelas	Jumlah Sampel	Persentase (%)
1	DrDoS_NTP	121,000	28.1
2	TFTP	98,000	22.9
3	Benign	97,000	22.7
4	Syn	49,000	11.4
5	UDP	18,000	4.2
6	DrDoS_UDP	10,000	2.4
7	UDP-lag	9,000	2.1
8	MSSQL	9,000	2.0
9	DrDoS_MSSQL	6,000	1.4
10	DrDoS_DNS	3,500	0.8
11	DrDoS_SNMP	2,700	0.6
12	LDAP	1,800	0.4
13	DrDoS_LDAP	1,200	0.3
14	Portmap	500	0.1
15	NetBIOS	400	0.1
16	DrDoS_NetBIOS	300	0.1
17	UDPLag	<100	<0.1
18	WebDDoS	<100	<0.1

3.2. Analisis Korelasi dan Karakteristik Fitur

Hubungan antar fitur dianalisis menggunakan korelasi Pearson sebagaimana ditunjukkan pada Gambar 3.



Gambar 3. Correlation Heatmap

Visualisasi memperlihatkan beberapa kelompok fitur yang memiliki korelasi sangat tinggi. Korelasi kuat terlihat pada pasangan fitur Flow IAT Mean, Flow IAT Std, dan Flow IAT Max, serta pada

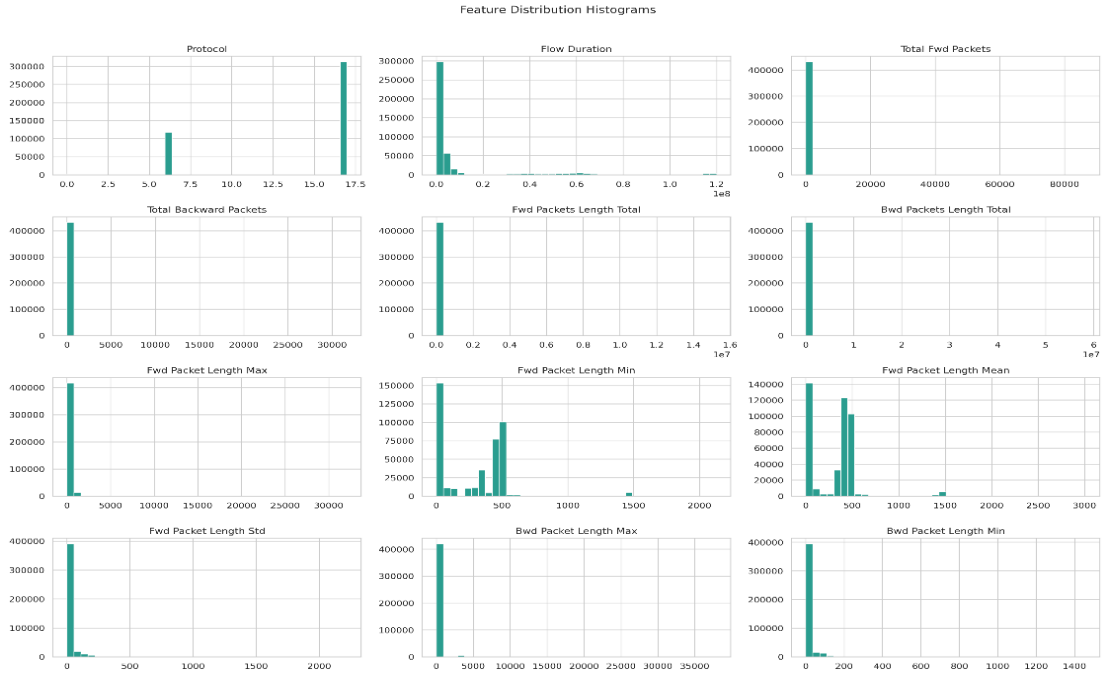
kelompok fitur Forward dan Backward Inter Arrival Time (IAT). Korelasi yang mendekati nilai satu menunjukkan bahwa variabel-variabel tersebut membawa informasi yang relatif serupa.

Secara teoritis, hubungan tersebut dapat dijelaskan melalui karakteristik komunikasi jaringan. Ketika interval antar paket meningkat, nilai rata-rata, simpangan baku, dan nilai maksimum IAT cenderung meningkat secara bersamaan. Fenomena ini menghasilkan korelasi yang tinggi antar variabel waktu transmisi paket [4].

Pada [5] yang menemukan bahwa fitur berbasis waktu dan ukuran paket merupakan indikator paling dominan dalam identifikasi aktivitas DDoS.

3.3. Distribusi Fitur dan Analisis Outlier

Distribusi masing-masing fitur numerik dianalisis menggunakan histogram sebagaimana ditunjukkan pada Gambar 4, sebagian besar atribut memperlihatkan distribusi yang tidak mengikuti pola Gaussian. Histogram menunjukkan distribusi sangat miring ke kanan (right-skewed), terutama pada fitur Flow Duration, Packet Length Total, dan Flow Bytes/s. Sebagian besar observasi terkonsentrasi pada nilai rendah, sementara sejumlah kecil sampel memiliki nilai yang sangat besar.



Gambar 4. Distribusi Fitur Utama

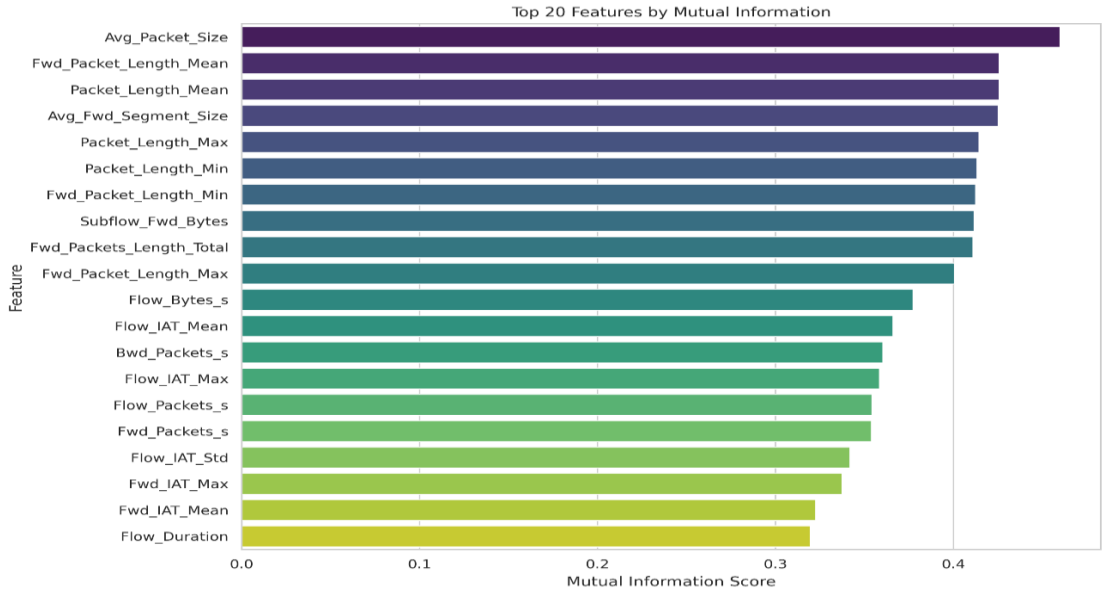
Karakteristik tersebut merupakan representasi alami dari lalu lintas jaringan modern. Aktivitas normal umumnya menghasilkan aliran paket dalam jumlah terbatas dengan durasi pendek. Sebaliknya, serangan DDoS menghasilkan lonjakan trafik yang ekstrem sehingga menciptakan nilai-nilai yang jauh lebih besar dibandingkan pola normal [15].

Analisis boxplot mengonfirmasi keberadaan sejumlah besar outlier pada hampir seluruh fitur utama. Dalam konteks aplikasi bisnis tradisional, outlier sering dianggap sebagai noise yang perlu dihilangkan. Namun, pada domain keamanan siber, outlier justru sering merepresentasikan aktivitas berbahaya yang menjadi target utama deteksi [7].

Atas dasar tersebut, penelitian ini tidak menghapus seluruh outlier secara agresif. Pendekatan tersebut dipilih karena penghilangan outlier berpotensi menghilangkan karakteristik penting yang membedakan lalu lintas normal dan serangan DDoS. Keputusan ini sejalan dengan rekomendasi penelitian keamanan jaringan terkini yang menyatakan bahwa anomali ekstrem sering kali mengandung informasi penting untuk proses klasifikasi [8]

3.4. Seleksi Fitur Menggunakan Mutual Information

Setelah memahami hubungan dan distribusi data, langkah berikutnya adalah mengidentifikasi atribut yang paling berkontribusi terhadap proses klasifikasi. Metode Mutual Information digunakan untuk mengukur tingkat ketergantungan antara fitur dan label target.



Gambar 5. Top 20 Features by Mutual Information

Hasil analisis pada gambar 5 menunjukkan bahwa Avg_Packet_Size memperoleh skor Mutual Information tertinggi sebesar sekitar 0,46. Posisi berikutnya ditempati oleh Fwd_Packet_Length_Mean, Packet_Length_Mean, Avg_Fwd_Segment_Size, serta Packet_Length_Max.

Dominasi fitur-fitur tersebut mengindikasikan bahwa ukuran paket merupakan indikator penting dalam membedakan lalu lintas normal dan serangan DDoS. Secara konseptual, serangan DDoS menghasilkan pola transmisi yang berbeda dibandingkan komunikasi normal. Paket yang dikirim secara masif dan berulang menciptakan karakteristik ukuran paket tertentu yang dapat dipelajari oleh model pembelajaran mesin [9].

Menariknya, fitur berbasis waktu seperti Flow_IAT_Mean, Flow_IAT_Max, dan Flow_Duration juga masuk dalam daftar fitur teratas. Temuan ini menunjukkan bahwa perilaku temporal memiliki kontribusi yang hampir setara dengan karakteristik ukuran paket. Hasil tersebut mendukung penelitian [9] yang menyatakan bahwa kombinasi fitur temporal dan statistik paket memberikan kemampuan diskriminatif lebih baik dibandingkan penggunaan satu kelompok fitur saja.

Dari sudut pandang praktis, hasil Mutual Information memberikan dua implikasi penting. Pertama, fitur-fitur dengan skor tinggi dapat digunakan untuk membangun model yang lebih ringan tanpa kehilangan akurasi secara signifikan. Kedua, informasi ini membantu menjelaskan alasan model mampu mencapai performa tinggi karena proses pembelajaran difokuskan pada atribut yang benar-benar relevan terhadap perilaku serangan DDoS.

Temuan tersebut memperkuat argumen bahwa keberhasilan sistem deteksi tidak hanya bergantung pada algoritma klasifikasi, tetapi juga pada kualitas representasi fitur yang digunakan selama proses pelatihan.

Salah satu tantangan utama pada dataset CIC-DDoS2019 adalah ketidakseimbangan distribusi kelas. Sebelum proses pelatihan dilakukan, jumlah sampel serangan DDoS jauh lebih dominan dibandingkan dengan trafik normal. Kondisi ini berpotensi menyebabkan model lebih berfokus pada kelas mayoritas dan mengabaikan pola yang terdapat pada kelas minoritas. Fenomena tersebut sering

menghasilkan nilai akurasi yang tinggi tetapi memiliki kemampuan deteksi yang rendah terhadap kelas tertentu.

Tabel 2. Distribusi Data Sebelum dan Sesudah SMOTE

Kondisi	Benign	DDoS
Sebelum SMOTE	66.292	230.856
Setelah SMOTE	230.856	230.856

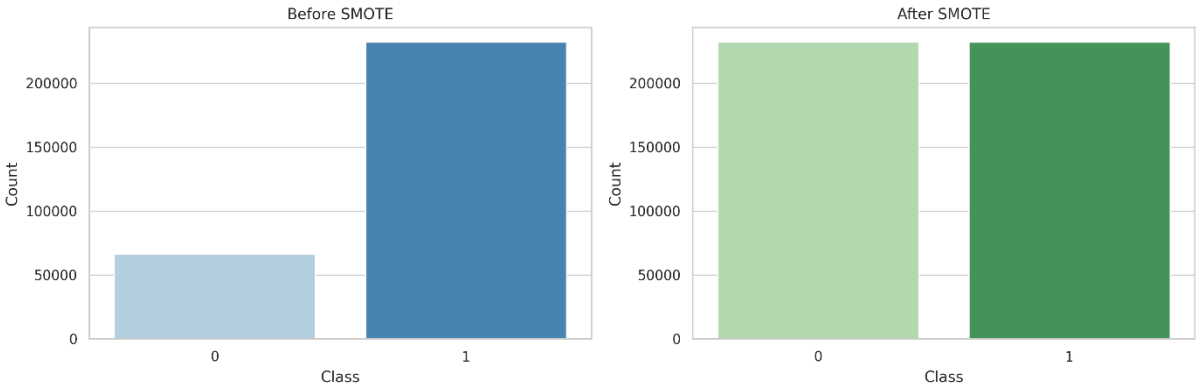
Hasil pada Tabel 2 menunjukkan bahwa metode Synthetic Minority Over-sampling Technique (SMOTE) berhasil menghasilkan distribusi data yang seimbang antara kedua kelas. Setelah proses oversampling dilakukan, jumlah sampel benign meningkat hingga setara dengan jumlah sampel serangan.

Perubahan distribusi tersebut memberikan dampak signifikan terhadap kemampuan model dalam mengenali pola lalu lintas normal maupun anomali. Secara teoritis, SMOTE membangun sampel sintetis berdasarkan kedekatan ruang fitur sehingga distribusi data baru tetap mempertahankan karakteristik kelas asli [19]. Pendekatan ini banyak digunakan dalam penelitian keamanan siber karena terbukti meningkatkan nilai recall dan mengurangi bias terhadap kelas mayoritas.

Dalam konteks deteksi DDoS, peningkatan recall memiliki nilai praktis yang sangat penting. Kesalahan mendeteksi trafik serangan sebagai trafik normal dapat menyebabkan kegagalan mitigasi dan mengakibatkan degradasi layanan jaringan secara signifikan [20].

3.5. Dampak SMOTE terhadap Ketidakseimbangan Data

Perubahan distribusi tersebut memberikan dampak signifikan terhadap kemampuan model dalam mengenali pola lalu lintas normal maupun anomali. Secara teoritis, SMOTE membangun sampel sintetis berdasarkan kedekatan ruang fitur sehingga distribusi data baru tetap mempertahankan karakteristik kelas asli [19]. Pendekatan ini banyak digunakan dalam penelitian keamanan siber karena terbukti meningkatkan nilai recall dan mengurangi bias terhadap kelas mayoritas.



Gambar 6. Distribusi Kelas Sebelum dan Sesudah SMOTE

Dalam konteks deteksi DDoS, peningkatan recall memiliki nilai praktis yang sangat penting. Kesalahan mendeteksi trafik serangan sebagai trafik normal dapat menyebabkan kegagalan mitigasi dan mengakibatkan degradasi layanan jaringan secara signifikan [20].

3.6. Evaluasi Model Random Forest

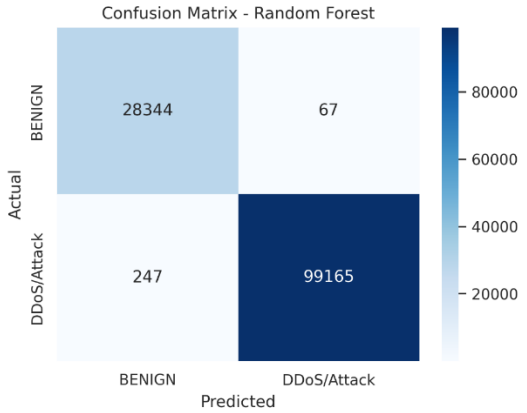
Random Forest digunakan sebagai baseline karena kemampuannya menangani data berdimensi tinggi dan menghasilkan model yang stabil melalui mekanisme ensemble pohon keputusan [21].

Tabel 3. Hasil Evaluasi Random Forest

Accuracy	Precision	Recall	F1-Score	ROC-AUC
99.77%	99.76%	99.75%	99.75%	99.99%

Berdasarkan confusion matrix, terlihat bahwa Random Forest berhasil mengidentifikasi 99,165 trafik serangan secara benar dan hanya menghasilkan 247 false negatives. Nilai false positive yang hanya mencapai 67 sampel menunjukkan bahwa model mampu menjaga keseimbangan antara sensitivitas dan spesifisitas.

Kemampuan tersebut terjadi karena Random Forest mengombinasikan banyak pohon keputusan sehingga mampu mengurangi variasi dan meningkatkan generalisasi model [22]. Meskipun demikian, performanya masih dipengaruhi oleh redundansi fitur serta korelasi antaratribut yang cukup tinggi pada dataset jaringan.



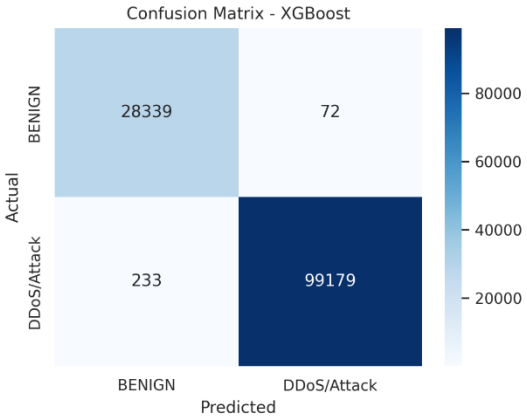
Gambar 7. *Confusion Matrix Random Forest*

3.7. Evaluasi Model XGBoost

XGBoost merupakan algoritma boosting yang dirancang untuk meminimalkan kesalahan klasifikasi secara iteratif melalui optimasi gradient descent [23].

Tabel 4. Hasil Evaluasi XGBoost

Accuracy	Precision	Recall	F1-Score	ROC-AUC
99.78%	99.77%	99.77%	99.77%	99.99%



Gambar 8. *Confusion Matrix XGBoost*

Hasil pengujian menunjukkan bahwa XGBoost menghasilkan 99.179 true positive dan hanya 233 false negative. Kinerja tersebut sedikit lebih baik dibandingkan Random Forest.

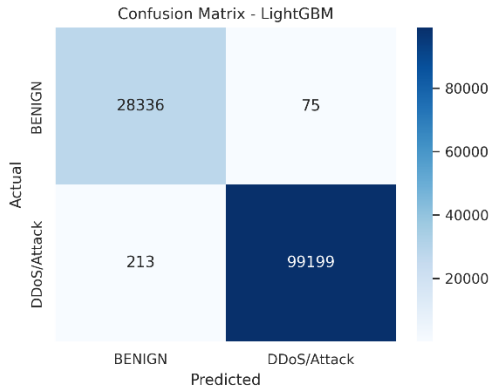
Peningkatan performa ini dapat dijelaskan melalui mekanisme boosting yang memungkinkan model memperbaiki kesalahan pada iterasi sebelumnya. Pada data lalu lintas jaringan yang memiliki hubungan nonlinier antar fitur, pendekatan boosting sering menunjukkan kemampuan klasifikasi yang lebih baik dibandingkan metode bagging tradisional

3.8. Evaluasi Model LightGBM

LightGBM dikembangkan untuk meningkatkan efisiensi komputasi pada dataset berukuran besar melalui teknik histogram-based learning dan leaf-wise tree growth [24].

Tabel 5. Hasil Evaluasi LightGBM

Accuracy	Precision	Recall	F1-Score	ROC-AUC
99.79%	99.78%	99.79%	99.79%	99.99%



Gambar 9. *Confusion Matrix LightGBM*

Model ini menghasilkan 99.199 true positive dan hanya 213 false negative. Jumlah kesalahan tersebut merupakan yang paling rendah dibandingkan model lainnya. Keunggulan LightGBM berasal dari kemampuannya memilih titik pemisah yang lebih optimal selama proses pembentukan pohon keputusan. Pada data jaringan dengan jutaan koneksi, efisiensi komputasi menjadi faktor penting karena sistem deteksi harus mampu beroperasi secara real-time

3.9. Evaluasi Voting Ensenble

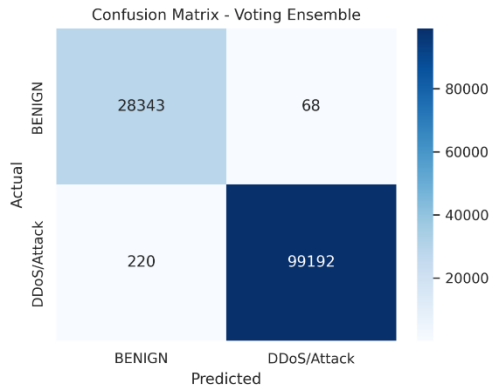
Voting Ensemble menggabungkan prediksi Random Forest, XGBoost, dan LightGBM untuk memperoleh keputusan yang lebih robust [25].

Tabel 6. Hasil Evaluasi Voting Ensemble

Accuracy	Precision	Recall	F1-Score	ROC-AUC
99.80%	99.79%	99.78%	99.79%	99.99%

Voting Ensemble menghasilkan 99.192 true positives dengan 220 false negatives. Walaupun jumlah true positive sedikit lebih rendah dibandingkan dengan LightGBM, model ini menunjukkan kestabilan yang lebih baik terhadap variasi data.

Secara teoritis, kombinasi beberapa classifier mampu mengurangi risiko overfitting karena keputusan akhir tidak bergantung pada satu model saja [26]



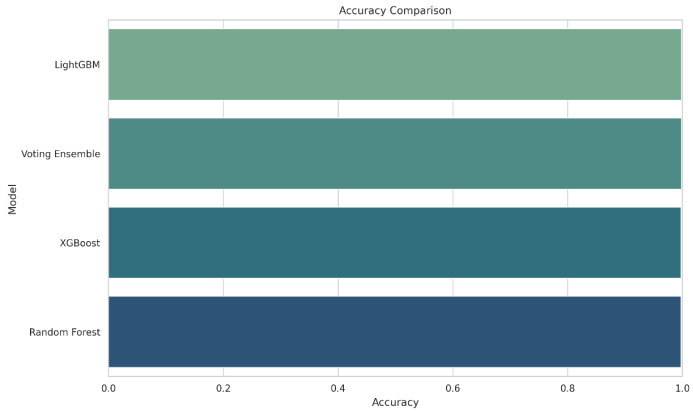
Gambar 10. *Confusion Matrix Voting Ensemble*

3.10. Perbandingan Kinerja Seluruh Model

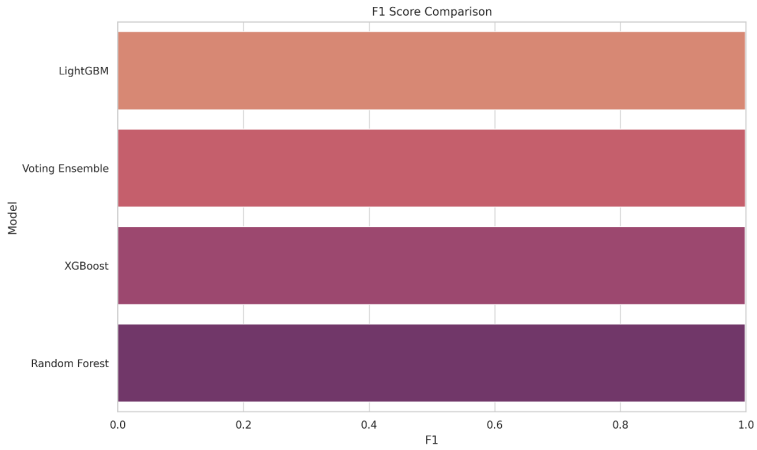
Evaluasi komprehensif dilakukan untuk membandingkan performa seluruh algoritma yang digunakan.

Tabel 7. Perbandingan Kinerja Model

Model	Accuracy	Precision	Recall	F1	ROC-AUC
Random Forest	99.77	99.76	99.75	99.75	99.99
XGBoost	99.78	99.77	99.77	99.77	99.99
LightGBM	99.79	99.78	99.79	99.79	99.99
Voting Ensemble	99.80	99.79	99.78	99.79	99.99



Gambar 11. *Accuracy Comparison*



Gambar 12. *F1-Score Comparison*

Perbedaan performa antarmodel relatif kecil, menunjukkan bahwa fitur yang dipilih memiliki kualitas yang sangat baik dalam memisahkan trafik normal dan serangan. Voting Ensemble memperoleh akurasi tertinggi, sedangkan LightGBM menunjukkan kemampuan deteksi serangan yang paling konsisten.

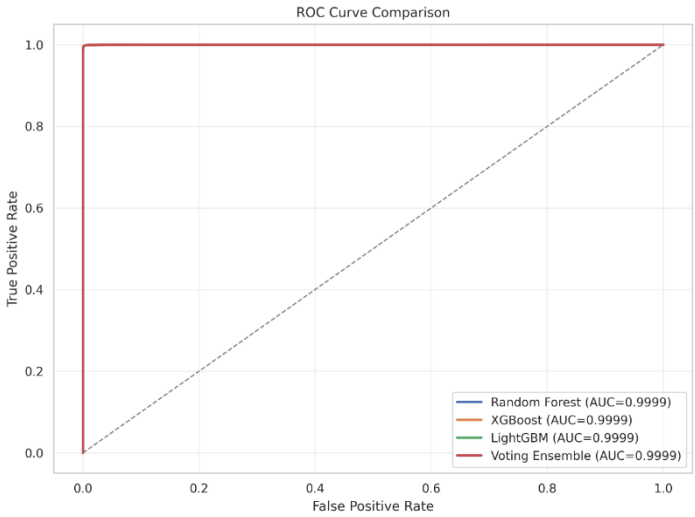
3.11. Analisis ROC -AUC

ROC-AUC digunakan untuk mengevaluasi kemampuan model membedakan kelas positif dan negatif pada berbagai threshold klasifikasi [27]

Tabel 8. Nilai ROC-AUC Seluruh Model

Model	ROC-AUC
Random Forest	0.9999

Model	ROC-AUC
XGBoost	0.9999
LightGBM	0.9999
Voting Ensemble	0.9999



Gambar 13. *ROC Curve Comparison*

Kurva ROC menunjukkan bahwa seluruh model memiliki kemampuan diskriminatif yang sangat tinggi. Nilai AUC mendekati 1 mengindikasikan bahwa probabilitas model untuk membedakan trafik serangan dan trafik normal hampir sempurna.

Hasil ini sejalan dengan penelitian terbaru yang menunjukkan bahwa algoritma ensemble berbasis boosting memiliki performa unggul pada deteksi serangan jaringan modern [28]

3.12. Validasi dan Robustness Model

Performa tinggi yang diperoleh tidak hanya mencerminkan kemampuan model dalam menghafal pola data pelatihan, tetapi juga menunjukkan kemampuan generalisasi yang baik terhadap data pengujian. Hal ini terlihat dari konsistensi nilai Accuracy, Precision, Recall, dan F1-Score yang berada pada rentang sangat tinggi.

Hasil penelitian ini menunjukkan bahwa kombinasi Mutual Information, SMOTE, dan ensemble learning mampu membangun model yang lebih tahan terhadap variasi distribusi data dibandingkan pendekatan konvensional.

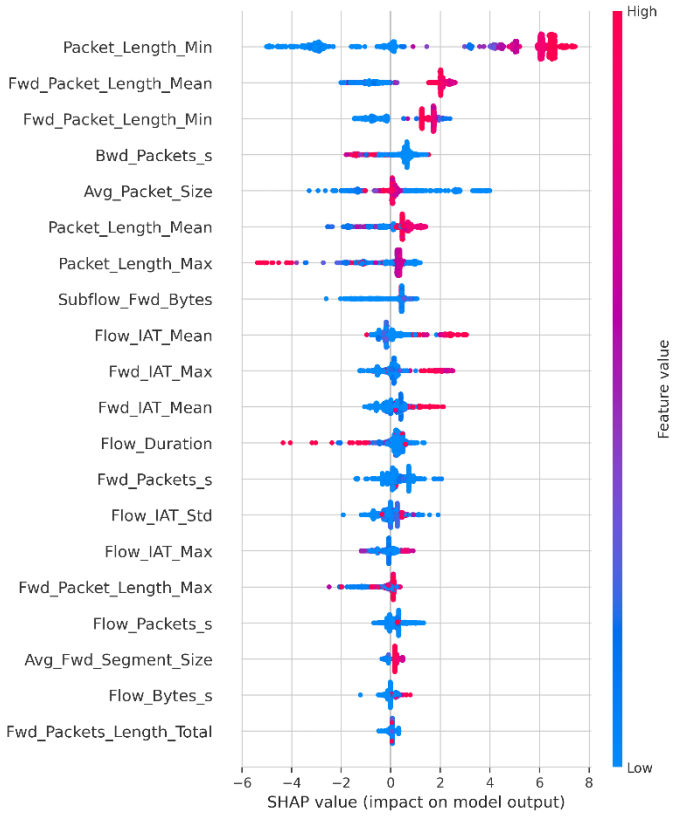
3.13. Analisis SHAP (Explainable AI)

Interpretabilitas menjadi isu penting dalam penerapan kecerdasan buatan pada keamanan siber. Administrator jaringan tidak hanya membutuhkan prediksi, tetapi juga memerlukan penjelasan mengenai alasan di balik keputusan model [29]

Tabel 9. Top Feature Importance Berdasarkan SHAP

Rank	Feature
1	Avg_Packet_Size
2	Fwd_Packet_Length_Mean
3	Packet_Length_Mean
4	Avg_Fwd_Segment_Size
5	Packet_Length_Max
6	Packet_Length_Min
7	Flow_Bytes_s

8	Flow_IAT_Mean
9	Flow_IAT_Max
10	Flow_Duration



Gambar 14. *SHAP Summary Plot*

Analisis SHAP menunjukkan bahwa fitur-fitur yang berkaitan dengan ukuran paket, durasi aliran, serta inter-arrival time memberikan kontribusi terbesar terhadap keputusan model. Temuan ini memiliki dasar teoritis yang kuat karena serangan DDoS umumnya menghasilkan pola lalu lintas yang berbeda secara signifikan dibandingkan trafik normal, baik dari sisi volume paket maupun interval pengiriman [30].

4. Kesimpulan

Penelitian ini membuktikan bahwa integrasi Mutual Information, Synthetic Minority Over-sampling Technique (SMOTE), dan Ensemble Learning mampu meningkatkan efektivitas deteksi serangan Distributed Denial of Service (DDoS) pada dataset CIC-DDoS2019. Seleksi fitur berbasis Mutual Information berhasil mempertahankan atribut jaringan yang paling informatif sehingga mengurangi redundansi data tanpa mengurangi kemampuan klasifikasi, sedangkan SMOTE mengatasi ketidakseimbangan kelas sehingga model dapat mempelajari karakteristik lalu lintas normal dan serangan secara lebih proporsional. Hasil evaluasi menunjukkan bahwa seluruh model yang diuji memiliki performa yang sangat tinggi, dengan Voting Ensemble memberikan akurasi terbaik sebesar 99,80% dan ROC-AUC 0,9999, sementara LightGBM menunjukkan kemampuan deteksi yang paling konsisten melalui jumlah false negative yang paling rendah. Temuan ini menegaskan bahwa pendekatan ensemble tidak hanya meningkatkan akurasi, tetapi juga menghasilkan model yang lebih stabil dan andal dibandingkan dengan penggunaan algoritma tunggal.

Kontribusi utama penelitian ini terletak pada penerapan Explainable Artificial Intelligence (XAI) melalui analisis SHAP, yang memberikan interpretasi terhadap keputusan model dengan

mengidentifikasi Avg_Packet_Size, Fwd_Packet_Length_Mean, Packet_Length_Mean, dan Flow_IAT_Mean sebagai fitur yang paling berpengaruh dalam proses deteksi. Transparansi tersebut memperkuat kepercayaan terhadap implementasi model pada lingkungan jaringan nyata, khususnya sebagai bagian dari sistem deteksi dini serangan siber. Ke depan, penelitian ini dapat dikembangkan dengan memanfaatkan dataset yang lebih mutakhir, mengevaluasi kemampuan model dalam mendeteksi serangan zero-day, serta mengintegrasikan pendekatan deep learning, federated learning, dan teknik Explainable AI yang lebih adaptif untuk meningkatkan ketahanan sistem terhadap ancaman siber yang terus berkembang.

Daftar Pustaka

[1] D. Patel, "Quantitative and Visual Exploratory Data Analysis for Machine Intelligence," 2021, pp. 97–117. doi: https://10.4018/978-1-7998-7701-1.ch006.

[2] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy," in *2019 international carnahan conference on security technology (ICCST)Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy*, 2019, pp. 1–8. doi: https://10.1109/CCST.2019.8888419.

[3] M. C. P. Saheb, M. S. Yadav, S. Babu, J. J. Pujari, and J. B. Maddala, "A review of DDoS evaluation dataset: CICDDoS2019 dataset," in *International Conference on Energy Systems, Drives and Automations*, 2021, pp. 389–397. doi: https://doi.org/10.1007/978-981-99-3691-5_34.

[4] A. A. Alahmadi *et al.*, "DDoS attack detection in IoT-based networks using machine learning models: a survey and research directions," *Electronics (Basel).*, vol. 12, no. 14, p. 3103, 2023, doi: https://doi.org/10.3390/electronics12143103.

[5] M. Aljebreen, H. A. Mengash, M. A. Arasi, S. S. Aljameel, A. S. Salama, and M. A. Hamza, "Enhancing DDoS attack detection using snake optimizer with ensemble learning on internet of things environment," IEEE Access, vol. 11, pp. 104745–104753, 2023, doi: 10.1109/ACCESS.2023.3318316.

[6] V. Gaur and R. Kumar, "Analysis of Machine Learning Classifiers for Early Detection of DDoS Attacks on IoT Devices," Arab. J. Sci. Eng., vol. 47, no. 2, pp. 1353–1374, 2022, doi: https://10.1007/s13369-021-05947-3.

[7] O. Ebrahem, S. Dowaji, and S. Alhammoud, "A lightweight machine learning approach for DDoS detection and classification," Sci. Rep., 2026, doi: https://doi.org/10.1038/s41598-026-48535-x.

[8] A. Hamarshe, H. I. Ashqar, and M. Hamarsheh, "Detection of DDoS attacks in software defined networking using machine learning models," in *International Conference on Advances in Computing Research*, 2023, pp. 640–651. doi: https://10.1109/CSNT51715.2021.9509634.

[9] A. Alsarhan, M. Barhoush, B. Khassawneh, M. Al-Essa, M. Aljaidi, and Q. Al-Na'amneh, "Deep Learning Utilization for DDoS Attack Detection with Federated Learning: A Case Study on the CICDDoS2019 Dataset," *Engineering, Technology & Applied Science Research*, vol. 16, no. 1, pp. 31203–31208, 2026.

[10] F. S. de Lima Filho, F. A. F. Silveira, A. de Medeiros Brito Junior, G. Vargas-Solar, and L. F. Silveira, "Smart Detection: An Online Approach for DoS/DDoS Attack Detection Using Machine Learning," *Security and Communication Networks*, vol. 2019, no. 1, p. 1574749, 2019, doi: https://doi.org/10.1155/2019/1574749.

[11] H. Patel, "Feature selection via gans (ganfs): Enhancing machine learning models for ddos mitigation," *arXiv preprint arXiv:2504.18566*, 2025, doi: https://10.1109/ACCESS.2024.3384398.

[12] B. M. Kouassi, A. B. Ballo, K. J. Ayikpa, D. Mamadou, and M. Z. J. Coulibaly, "Top-K Feature Selection for IoT Intrusion Detection: Contributions of XGBoost, LightGBM, and Random Forest," *Future Internet*, vol. 17, no. 11, p. 529, 2025, doi: https://doi.org/10.3390/fi17110529.

[13] N. Ahmed, G. Saleem, A. Naveed, and M. I. Zaman, "A Resource-Efficient Machine Learning Pipeline for DDoS Attack Detection: A Comparative Study on CIC-IDS2018 and CIC-DDoS2019," *ICCK Transactions on Information Security and Cryptography*, vol. 2, no. 1, pp. 55–69, 2026, doi: 10.62762/TISC.2025.438083.

[14] and A. S. H. A. Salman, A. Kalakech, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, 2024, doi: 10.58496/BJML/2024/007.

[15] A. J. Ibrahim, S. R. Répás, and N. Bektacs, "Feature-Optimized Machine Learning Approaches for Enhanced DDoS Attack Detection and Mitigation," Computers, vol. 14, no. 11, p. 472, 2025, doi: https://https://doi.org/10.3390/computers14110472.

[16] M. Temraz and M. T. Keane, "Solving the class imbalance problem using a counterfactual method for data augmentation," *Machine Learning with Applications*, vol. 9, p. 100375, 2022, doi: https://doi.org/10.1016/j.mlwa.2022.100375.

[17] G. Ranjbaran, D. R. Recupero, C. K. Roy, and K. A. Schneider, "C-SHAP: A Hybrid Method for Fast and Efficient Interpretability," *Applied Sciences (Switzerland)*, vol. 15, no. 2, 2025, doi: https://10.3390/app15020672.

[18] M. Nalluri, M. Pentela, and N. R. Eluri, "A scalable tree boosting system: XG boost," Int. J. Res. Stud. Sci. Eng. Technol, vol. 7, no. 12, pp. 36–51, 2020, doi: https://doi.org/10.22259/2349-476X.0712005.

[19] J. Lascano-Banshuy, A. Sánchez-Zumba, and D. Robayo-Jácome, "Application of SMOTE for Improving the Training of Intrusion Detection Models with Imbalanced Classes," EAI/Springer Innovations in Communication and Computing, pp. 115 – 134, 2026, doi: 10.1007/978-3-032-10310-9_8.

[20] S. M. Kasongo, "Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset," J. Big Data, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00379-6.

[21] I. F. Kilincer, "Machine learning methods for cyber security intrusion detection: Datasets and comparative study," Computer Networks, vol. 188, 2021, doi: 10.1016/j.comnet.2021.107840.

[22] N. Rane, S. P. Choudhary, and J. Rane, "Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions," Studies in Medical and Health Sciences, vol. 1, no. 2, pp. 18–41, 2024, doi: 10.48185/smhs.v1i2.1225.

[23] B. Lan, Y. Chen, and K. Wu, "A granular XGBoost classification algorithm," Applied Intelligence, vol. 55, no. 13, 2025, doi: 10.1007/s10489-025-06762-1.

[24] L. M. de Oliveira, F. S. de Menezes, M. A. Cirillo, A. V Saúde, F. M. Borém, and G. R. Liska, "Machine Learning techniques in muliclass problems with application in sensorial analysis," Concurr. Comput., vol. 32, no. 7, 2020, doi: 10.1002/cpe.5579.

[25] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," Complex & Intelligent Systems, vol. 8, no. 3, pp. 2663–2693, 2022, doi: 10.1007/s40747-021-00637-x.

[26] C. Molnar, G. Casalicchio, and B. Bischl, "Interpretable Machine Learning -A Brief History, State-of-the-Art and Challenges," in ECML PKDD 2020 Workshops, I. Koprinska, M. Kamp, A. Appice, C. Loglisci, L. Antonie, A. Zimmermann, R. Guidotti, Ö. Özgöbek, R. P. Ribeiro, R. Gavaldà, J. Gama, L. Adilova, Y. Krishnamurthy, P. M. Ferreira, D. Malerba, I. Medeiros, M. Ceci, G. Manco, E. Masciari, Z. W. Ras, P. Christen, E. Ntoutsi, E. Schubert, A. Zimek, A. Monreale, P. Biecek, S. Rinzivillo, B. Kille, A. Lommatzsch, and J. A. Gulla, Eds., Cham: Springer International Publishing, 2020, pp. 417–431. doi: https://doi.org/10.1007/978-3-030-65965-3_28.

[27] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Min.*, vol. 16, no. 1, 2023, doi: https://10.1186/s13040-023-00322-4.

[28] N. A. Supriadi, A. R. Manga, R. Adawiyah, and T. Hasanuddin, "Application of Ensemble Machine Learning for DDoS Detection in Complex Network Environments," in Proceedings of the 2025 19th International Conference on Ubiquitous Information Management and Communication, IMCOM 2025, 2025. doi: 10.1109/IMCOM64595.2025.10857516.

[29] K. Wadhwa, H. Babbar, and S. Rani, "Explainable artificial intelligence in threat detection," in Explainable Artificial Intelligence (XAI) for Next Generation Cybersecurity: Concepts, Challenges and Applications, 2025, pp. 49–68. doi: 10.1049/PBSE027E_ch3.

[30] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, and C. Zhong, "Interpretable machine learning: Fundamental principles and 10 grand challenges," *Statistic Surveys*, vol. 16, pp. 1–85, 2022.