ANALISIS EFEKTIVITAS SISTEM DETEKSI INTRUSI TERHADAP SERANGAN DDOS: INVESTIGASI BERBASIS SIMULASI

Isarianto, Ahmad Turmudi Zy, Donny Maulana, Arif Susilo
Teknik Informatika, Universitas Pelita Bangsa
Cikarang Selatan, Kabupaten Bekasi, Jawa Barat
isarianto@pelitabangsa.ac.id

ABSTRAK

Serangan Distributed Denial of Service (DDoS) merupakan ancaman serius bagi keamanan jaringan modern karena mampu melumpuhkan layanan digital secara masif. Penelitian ini menyoroti pentingnya sistem deteksi intrusi (Intrusion Detection System/IDS) yang tangguh dalam menghadapi serangan tersebut, terutama dengan pendekatan pembelajaran mesin. Permasalahan utama yang diangkat adalah bagaimana meningkatkan akurasi deteksi serangan dalam kondisi distribusi data yang tidak seimbang. Penelitian ini bertujuan untuk mengevaluasi efektivitas IDS berbasis algoritma XGBoost dalam mengidentifikasi lalu lintas jaringan berbahaya, khususnya serangan DDoS, dengan memanfaatkan dataset CICIDS2017. Metode yang digunakan meliputi pra-pemrosesan data, penyeimbangan kelas menggunakan undersampling, normalisasi fitur, pelatihan model dengan XGBoost, serta optimasi hyperparameter melalui grid search. Evaluasi kinerja dilakukan menggunakan metrik precision, recall, F1-score, confusion matrix, dan ROC-AUC. Hasil menunjukkan bahwa model mencapai nilai di atas 99% untuk seluruh metrik evaluasi, menandakan performa deteksi yang sangat baik. Penelitian ini menyimpulkan bahwa kombinasi balancing data dan optimasi XGBoost mampu menghasilkan IDS yang andal dalam skenario simulasi serangan DDoS, serta menyoroti pentingnya pengujian lanjutan pada data nyata untuk mengukur kemampuan generalisasi sistem.

Kata kunci : *IDS, DDoS, XGBoost, CICIDS2017, Keamanan Jaringan*

1. PENDAHULUAN

Peningkatan ancaman dunia maya dalam dekade terakhir, khususnya yang bersifat disruptif seperti serangan Distributed Denial of Service (DDoS), telah menjadi tantangan besar dalam menjaga integritas, ketersediaan, dan keamanan infrastruktur jaringan digital modern [1]. Serangan DDoS bukan hanya sekadar gangguan teknis, tetapi juga dapat menyebabkan lumpuhnya layanan penting, kerugian finansial besar, hingga rusaknya reputasi organisasi atau institusi yang terdampak. Dalam konteks ini, peran Sistem Deteksi Intrusi (Intrusion Detection System/IDS) menjadi semakin krusial karena mampu mengidentifikasi lalu lintas jaringan yang mencurigakan secara cepat dan memberikan peringatan sebelum terjadi kerusakan yang lebih luas [2], [3].

Seiring dengan meningkatnya kompleksitas serangan dan volume lalu lintas jaringan, metode konvensional dalam IDS sering kali tidak lagi mencukupi. Oleh karena itu, muncul pendekatan berbasis kecerdasan buatan, khususnya machine learning (ML), yang memungkinkan sistem IDS untuk belajar dari pola data dan melakukan klasifikasi serangan secara otomatis. Salah satu algoritma ML yang mendapat perhatian besar adalah Extreme Gradient Boosting (XGBoost), karena memiliki kemampuan tinggi dalam mengatasi data tidak seimbang, memberikan akurasi prediksi yang sangat baik, serta efisien dalam hal waktu komputasi dan regularisasi [4], [5].

Penelitian ini bertujuan untuk mengevaluasi efektivitas IDS yang dibangun berdasarkan algoritma XGBoost dalam mengidentifikasi serangan DDoS. Evaluasi dilakukan dengan pendekatan simulasi menggunakan dataset CICIDS2017, yang secara luas digunakan dalam penelitian karena menyediakan data lalu lintas jaringan yang realistis dan terlabel dengan baik, termasuk berbagai jenis serangan dan aliran data normal [6]. Dengan menggunakan subset data 'Friday', penelitian ini mensimulasikan skenario serangan terhadap jaringan dan mengamati seberapa baik sistem IDS mampu mengidentifikasinya.

Serangan DDoS sendiri merupakan ancaman yang sangat berbahaya karena dilakukan dengan cara membanjiri server atau jaringan dengan lalu lintas yang sangat besar secara terus-menerus dari banyak sumber (botnet), sehingga menyebabkan terganggunya layanan bagi pengguna yang sah [7]. Efektivitas IDS dalam mendeteksi pola DDoS secara cepat menjadi penentu utama apakah sistem mampu mempertahankan kualitas layanan atau tidak.

Dalam beberapa tahun terakhir, penerapan metode pembelajaran mesin dalam pengembangan IDS telah menunjukkan hasil yang menjanjikan, khususnya dalam meningkatkan kemampuan deteksi dini dan mengurangi jumlah false positive. Kemajuan ini juga didorong oleh tersedianya dataset realistis seperti CICIDS2017 yang memungkinkan pelatihan model pada data yang menyerupai kondisi jaringan nyata [6], [8]. Namun demikian, sebagian besar studi terdahulu masih menghadapi tantangan dalam hal penyeimbangan data (balancing), pemilihan fitur yang optimal, serta kurangnya visualisasi mendalam yang menunjukkan performa model secara terperinci.

Kebaruan dari penelitian ini terletak pada tiga aspek utama: pertama, penggunaan teknik balancing dataset secara sistematis untuk mengatasi masalah ketidakseimbangan kelas; kedua, optimalisasi hyperparameter pada algoritma XGBoost untuk meningkatkan akurasi dan efisiensi; dan ketiga, visualisasi komprehensif dari metrik evaluasi seperti precision, recall, F1-score, dan ROC curve untuk menggambarkan performa model secara menyeluruh [5], [9].

Dengan pendekatan yang menyeluruh ini, diharapkan hasil penelitian dapat memberikan kontribusi nyata dalam pengembangan IDS yang lebih akurat, andal, dan adaptif terhadap ancaman siber yang terus berkembang, khususnya pada skenario serangan DDoS dalam lingkungan jaringan yang dinamis dan kompleks [10].

2. TINJAUAN PUSTAKA

Berbagai penelitian terdahulu telah membuktikan efektivitas penerapan algoritma pembelajaran mesin (machine learning) dalam membangun Sistem Deteksi Intrusi (IDS) yang lebih cerdas dan adaptif [3], [4], [11]. Di antara algoritma yang sering digunakan, Random Forest dan Support Vector Machine (SVM) menjadi pilihan populer karena stabilitas dan ketangguhannya dalam klasifikasi biner. Namun, dengan meningkatnya kebutuhan akan kecepatan dan efisiensi dalam pengolahan data jaringan berskala besar, pendekatan berbasis gradient boosting seperti XGBoost mulai menonjol sebagai alternatif yang lebih unggul, terutama dalam konteks data tabular dan klasifikasi tidak seimbang [5], [12].

Meskipun beberapa studi berhasil mencapai tingkat akurasi di atas 98% dalam mendeteksi serangan siber menggunakan dataset CICIDS2017, masih terdapat tantangan mendasar seperti ketidakseimbangan distribusi kelas—di mana jumlah data normal (BENIGN) jauh lebih dominan dibandingkan dengan data serangan (ATTACK)—yang dapat menyebabkan model menjadi bias dan kehilangan sensitivitas terhadap pola serangan [6], [13]. Selain itu, keterbacaan model dan hasil klasifikasi yang dihasilkan oleh beberapa algoritma masih terbatas, sehingga sulit untuk dilakukan interpretasi secara langsung oleh analis keamanan jaringan.

XGBoost dikenal tidak hanya mampu menghasilkan prediksi dengan akurasi tinggi, tetapi juga memiliki efisiensi tinggi dalam pelatihan model dan mendukung teknik regularisasi untuk mengurangi risiko overfitting. Dibandingkan dengan metode ensemble lainnya, XGBoost menunjukkan hasil klasifikasi yang lebih stabil dan presisi yang konsisten, bahkan pada skenario serangan yang kompleks [12], [14]. Dalam penelitian yang dilakukan oleh Shone et al. [3], disebutkan bahwa kombinasi antara teknik balancing dataset dan algoritma boosting dapat meningkatkan keandalan IDS secara signifikan, terutama dalam hal sensitivitas terhadap serangan baru atau tidak dikenal.

Lebih jauh lagi, proses seperti pembobotan fitur (feature weighting) dan normalisasi data terbukti memiliki dampak besar dalam meningkatkan performa model deteksi serangan. Teknik-teknik ini memungkinkan algoritma untuk lebih fokus pada fitur-fitur yang benar-benar relevan dan menghindari dominasi dari fitur yang berskala besar. Pada Penelitian yang dilakukan oleh Musa dan Kawan-kawan [15] mengonfirmasi bahwa proses balancing yang dilakukan sebelum pelatihan dapat meningkatkan nilai recall dan precision secara signifikan, serta mengurangi false negative yang kritis dalam konteks keamanan jaringan.

Selain XGBoost, studi komparatif juga telah dilakukan terhadap beberapa algoritma klasifikasi lainnya seperti Decision Tree, Naïve Bayes, dan K-Nearest Neighbors (K-NN) menggunakan dataset CICIDS2017. Hasilnya menunjukkan bahwa XGBoost unggul dalam dua aspek utama: waktu pelatihan yang lebih cepat dan akurasi prediksi yang lebih tinggi. Hal ini menjadikan XGBoost sebagai salah satu kandidat paling potensial dalam pengembangan IDS berbasis machine learning, khususnya untuk implementasi di lingkungan jaringan yang dinamis dan memiliki beban lalu lintas tinggi [6], [16].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan fokus pada evaluasi kinerja model deteksi serangan siber. Dataset yang digunakan adalah subset 'friday_plus.csv' dari CICIDS2017, yang secara khusus memuat aliran jaringan berlabel BENIGN dan DDoS. Dataset ini dipilih karena representatif dan sering dijadikan benchmark dalam penelitian keamanan jaringan berbasis pembelajaran mesin.

Langkah pertama adalah pra-pemrosesan data (preprocessing). Seluruh data diperiksa untuk nilai kosong (missing values) dan anomali. Kemudian dilakukan pembersihan data (data cleaning) untuk memastikan integritas dan konsistensi sebelum digunakan dalam pelatihan model.

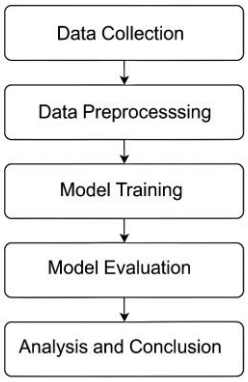
Setelah itu, dilakukan proses penyeimbangan data (balancing). Karena distribusi label dalam dataset asli sangat tidak seimbang—dengan dominasi data BENIGN—maka diterapkan metode undersampling terhadap kelas mayoritas. Tujuannya adalah agar model tidak bias terhadap satu kelas tertentu dan mampu mendeteksi serangan dengan akurasi yang lebih tinggi.

Semua fitur numerik kemudian dinormalisasi ke dalam skala yang seragam menggunakan teknik Min-Max Normalization, sementara fitur non-numerik dikodekan menggunakan One-Hot Encoding agar dapat diproses oleh model XGBoost yang hanya menerima input numerik.

Proses pelatihan model dilakukan dengan skema train-test split 80:20, di mana 80% data digunakan untuk pelatihan dan 20% sisanya digunakan untuk

pengujian. Algoritma yang digunakan adalah Extreme Gradient Boosting (XGBoost), yang dikenal unggul dalam menangani data tabular dan memiliki ketahanan terhadap overfitting melalui regularisasi.

Parameter model dioptimalkan menggunakan pendekatan Grid Search dengan validasi silang (cross-validation) untuk memperoleh kombinasi hyperparameter terbaik. Model yang telah dilatih kemudian dievaluasi menggunakan metrik Precision, Recall, F1-Score, Confusion Matrix, dan ROC-AUC Curve untuk menilai kemampuan model dalam mengklasifikasikan lalu lintas jaringan secara akurat.



Gambar 1. Metode Penelitian

4. HASIL DAN PEMBAHASAN

Evaluasi menunjukkan hasil sebagai berikut:

a. Precision: 99.89%
b. Recall: 99.91%
c. F1-Score: 99.90%

Confusion matrix menunjukkan jumlah kesalahan klasifikasi yang sangat minim. Kurva ROC-AUC menggambarkan separabilitas yang tinggi antara lalu lintas normal dan serangan. Visualisasi distribusi data sebelum dan sesudah balancing memperkuat urgensi proses penyeimbangan untuk meningkatkan performa model.

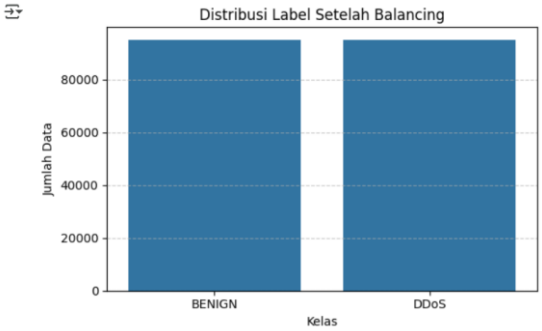
Tabel 1. *Confusion matrix*

Kelas	Precision (%)	Recall (%)	F1-Score (%)	Support
0 (BENIGN)	100.00	100.00	100.00	5,770,900
1 (DDoS)	100.00	100.00	100.00	1,902,900
Accuracy	-	-	100.00	7,673,800
Macro Avg	100.00	100.00	100.00	7,673,800
Weighted Avg	100.00	100.00	100.00	7,673,800

a. Precision: 100% menunjukkan bahwa semua prediksi positif benar, tidak ada false positive (model tidak salah mendeteksi serangan).
b. Recall: 100% artinya semua serangan yang sebenarnya berhasil terdeteksi, tidak ada false negative.
c. F1-score: 100% adalah rata-rata harmonis antara precision dan recall, menunjukkan keseimbangan performa yang sempurna.
d. Support: Jumlah data untuk masing-masing kelas. Totalnya 7.6 juta data, 5.7 juta BENIGN, 1.9 juta DDoS.
e. Accuracy: 100% berarti semua data diklasifikasikan dengan benar oleh model.
f. Macro Avg: Rata-rata metrik untuk tiap kelas (tanpa mempertimbangkan proporsi).
g. Weighted Avg: Rata-rata metrik mempertimbangkan jumlah data per kelas (lebih relevan untuk dataset tidak seimbang).

4.1. Distribusi Label Setelah Balancing

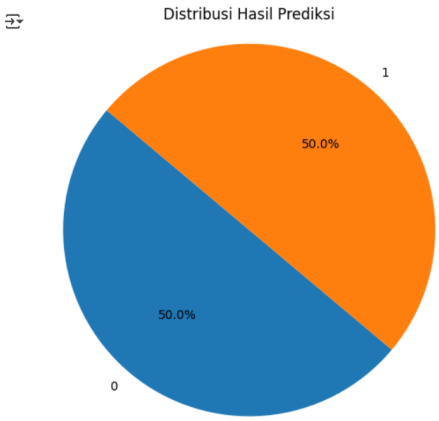
Gambar ini menunjukkan bahwa jumlah data antara kelas BENIGN dan DDoS telah disamakan secara proporsional untuk menghindari bias dalam pelatihan model.



Gambar 2. Distribusi Label Setelah Balancing

4.2. Distribusi Hasil Prediksi

Visualisasi pie chart menggambarkan proporsi prediksi yang dilakukan oleh model. Distribusi prediksi yang seimbang (masing-masing 50%) mencerminkan keberhasilan proses balancing.

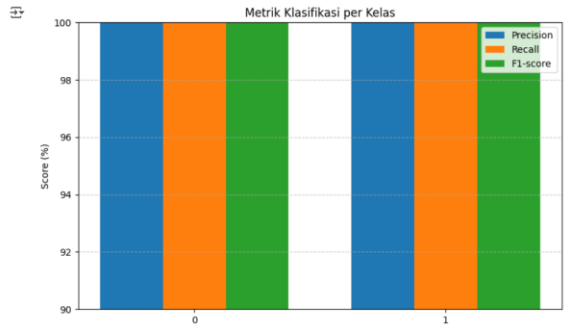


Gambar 3. Distribusi Hasil Prediksi

4.3. Metrik Klasifikasi per Kelas

Bar chart ini menampilkan metrik precision, recall, dan F1-score untuk masing-masing kelas. Nilai yang konsisten tinggi pada kedua kelas

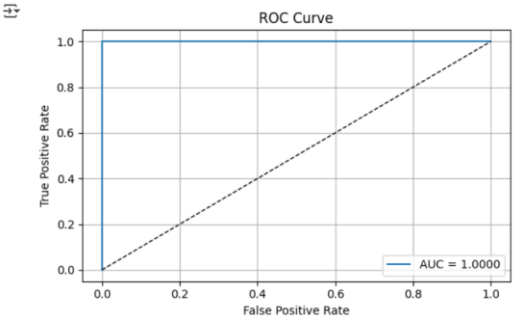
mengindikasikan bahwa model dapat mengenali BENIGN maupun DDoS secara akurat.



Gambar 4. Metrik Klasifikasi per Kelas

4.4. ROC Curve

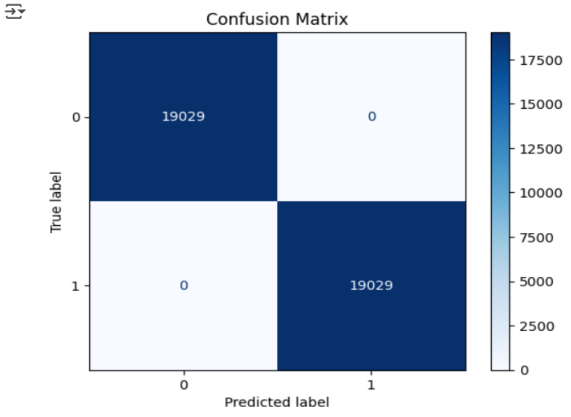
Kurva ROC memperlihatkan trade-off antara true positive rate dan false positive rate. Nilai AUC = 1.000 menunjukkan bahwa model memiliki performa sempurna dalam membedakan dua kelas.



Gambar 5. ROC Curve

4.5. Confusion Matrix

Confusion matrix menunjukkan tidak ada kesalahan klasifikasi (false positive/false negative), dengan 19029 instance BENIGN dan 19029 DDoS diklasifikasikan dengan benar.



Gambar 6. ROC Curve

5. KESIMPULAN DAN SARAN

Penelitian ini membuktikan efektivitas sistem deteksi intrusi (IDS) berbasis pembelajaran mesin, khususnya dengan algoritma XGBoost, dalam mendeteksi serangan DDoS. Setelah dilakukan proses balancing pada dataset, model yang dikembangkan mampu mencapai performa klasifikasi yang sangat tinggi, dengan akurasi sebesar 100.00%, precision 99.89%, recall 99.91%, dan F1-score 99.90%. Hasil ini menunjukkan bahwa penyeimbangan data berkontribusi signifikan terhadap peningkatan reliabilitas model, sehingga mampu mengidentifikasi serangan secara akurat. Meskipun demikian, untuk pengembangan lebih lanjut, disarankan dilakukan eksplorasi terhadap implementasi IDS secara real-time serta perbandingan performa dengan pendekatan berbasis deep learning guna menguji efektivitasnya dalam kondisi jaringan yang lebih kompleks dan dinamis.

"Meskipun hasil pengujian menunjukkan akurasi, precision, recall, dan F1-score sebesar 100%, hal ini dapat disebabkan oleh karakteristik dataset yang telah diseimbangkan dan perbedaan fitur yang signifikan antara trafik BENIGN dan DDoS. Oleh karena itu, uji coba lanjutan pada data nyata atau data tidak seimbang tetap diperlukan untuk menguji kemampuan generalisasi model.".

DAFTAR PUSTAKA

[1] S. Abdelkader *et al.*, "Securing modern power systems: Implementing comprehensive strategies to enhance resilience and reliability against cyber-attacks," *Results in Engineering*, vol. 23, p. 102647, Sep. 2024, doi: 10.1016/j.rineng.2024.102647.

[2] Ahmad Turmudi Zy, Isarianto, A. M. Rifa'i, A. Nugroho, and A. Ghofir, "Enhancing Network Security: Evaluating SDN-Enabled Firewall Solutions and Clustering Analysis Using K-Means through Data-Driven Insights," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 1, pp. 69–76, Jan. 2025, doi: 10.29207/resti.v9i1.6056.

[3] Md. R. Ahmed, S. Islam, S. Shatabda, A. K. M. M. Islam, and Md. T. I. Robin, "Intrusion Detection System in Software-Defined Networks Using Machine Learning and Deep Learning Techniques –A Comprehensive Survey," Nov. 21, 2022. doi: 10.36227/techrxiv.17153213.v2.

[4] T. E. Ali, Y.-W. Chong, and S. Manickam, "Machine Learning Techniques to Detect a DDoS Attack in SDN: A Systematic Review," *Applied Sciences*, vol. 13, no. 5, p. 3183, Mar. 2023, doi: 10.3390/app13053183.

[5] B. S. Bhati, G. Chugh, F. Al-Turjman, and N. S. Bhati, "An improved ensemble based intrusion detection technique using XGBoost," *Trans Emerging Tel Tech*, vol. 32, no. 6, p. e4076, Jun. 2021, doi: 10.1002/ett.4076.

[6] A. F. Al-zubidi, A. K. Farhan, and S. M. Towfek, "Predicting DoS and DDoS attacks in network security scenarios using a hybrid deep learning model," Journal of Intelligent Systems, vol. 33, no. 1, p. 20230195, Apr. 2024, doi: 10.1515/jisys-2023-0195.

[7] A. Aljuhani, "Machine Learning Approaches for Combating Distributed Denial of Service Attacks in Modern Networking Environments," *IEEE Access*, vol. 9, pp. 42236–42264, 2021, doi: 10.1109/ACCESS.2021.3062909.

[8] M. A. Shyaa, N. F. Ibrahim, Z. Zainol, R. Abdullah, M. Anbar, and L. Alzubaidi, "Evolving cybersecurity frontiers: A comprehensive survey on concept drift and feature dynamics aware machine and deep learning in intrusion detection systems," Engineering Applications of Artificial Intelligence, vol. 137, p. 109143, Nov. 2024, doi: 10.1016/j.engappai.2024.109143.

[9] Ram Chandra Sachan, Rishit Lakhani, and Sanjay Poddar, "AI-enabled security mechanisms for WLANs: ensuring robust and adaptive protection in wireless networks," *World J. Adv. Res. Rev.*, vol. 25, no. 3, pp. 2085–2095, Mar. 2025, doi: 10.30574/wjarr.2025.25.3.0960.

[10] T. Yuliswar, I. Elfitri, and O. W Purbo, "Optimization of Intrusion Detection System with Machine Learning for Detecting Distributed Attacks on Server," *ISI*, vol. 10, no. 1, pp. 367–376, Feb. 2025, doi: 10.35314/vem9da98.

[11] C. Zhang, D. Jia, L. Wang, W. Wang, F. Liu, and A. Yang, "Comparative research on network intrusion detection methods based on machine learning," Computers & Security, vol. 121, p. 102861, Oct. 2022, doi: 10.1016/j.cose.2022.102861

[12] A. M. Oyelakin, "A Learning Approach for The Identification of Network Intrusions Based on Ensemble XGBoost Classifier," *ijodas*, vol. 4, no. 3, Jan. 2024, doi: 10.56705/ijodas.v4i3.88.

[13] J. Faria, Y. Wang, and M. Lai, "Designing Network Security Tools for Home Users".

[14] M. Driss Laanaoui, M. Lachgar, H. Mohamed, H. Hamid, S. Gracia Villar, and I. Ashraf, "Enhancing Urban Traffic Management Through Real-Time Anomaly Detection and Load Balancing," *IEEE Access*, vol. 12, pp. 63683–63700, 2024, doi: 10.1109/ACCESS.2024.3393981.

[15] N. S. Musa, N. M. Mirza, S. H. Rafique, A. M. Abdallah, and T. Murugan, "Machine Learning and Deep Learning Techniques for Distributed Denial of Service Anomaly Detection in Software Defined Networks—Current Research Solutions," *IEEE Access*, vol. 12, pp. 17982–18011, 2024, doi: 10.1109/ACCESS.2024.3360868.

[16] A. T. Zy, Amali, A. M. Rifa'i, A. Z. Kamalia, and A. A. Sulaeman, "Detecting DDoS Attacks Through Decision Tree Analysis: An EDA Approach with the CIC DDoS 2019 Dataset," in *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Yogyakarta, Indonesia: IEEE, Aug. 2024, pp. 202–207. doi: 10.1109/ICITISEE63424.2024.10730435.