Vol. 6, No. 3, Desember 2025, hal. 530-535

Jurnal Computer Science and Information Technology (CoSciTech)

p-ISSN: 2723-567X
e-ISSN: 2723-5661

http://ejurnal.umri.ac.id/index.php/coscitech/index

Klasifikasi serangan DDoS dengan metode random forest dan teknik class weight pada dataset CICDDoS2019

Desti Mualfah*[1], Rudi Ardiansyah[2], Rahmad Gunawan[3]
Email: [1]destimualfah@umri.ac.id, [2]210401147@student.umri.ac.id, [3]goengoen78@umri.ac.id

[123]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau

Diterima: 12 Desember 2025 | Direvisi: 16 Desember 2025 | Disetujui: 20 Desember 2025

Abstrak

Perkembangan teknologi informasi yang semakin pesat membawa dampak signifikan terhadap berbagai aspek kehidupan, termasuk meningkatnya ketergantungan pada layanan berbasis jaringan. Namun, kondisi ini sekaligus memicu munculnya ancaman keamanan siber yang semakin kompleks, salah satunya serangan berupa *Distributed Denial of Service* (DDoS). Serangan ini berpotensi melumpuhkan ketersediaan layanan dengan membanjiri sistem target menggunakan lalu lintas data berlebih. Tantangan utama dalam mendeteksi DDoS adalah banyaknya variasi serangan serta ketidakseimbangan kelas pada data lalu lintas jaringan. Untuk mengatasi permasalahan tersebut, diperlukan pendekatan berbasis pembelajaran mesin yang mampu menangani kompleksitas pola serangan sekaligus menyeimbangkan distribusi data. Salah satu solusi dalam mengklasifikasi serangan DDoS adalah dengan memanfaatkan algoritma *Random Forest* dan teknik *Class Weight* pada dataset CICDDoS2019. Proses penelitian meliputi tahap pengumpulan dan eksplorasi data, pra-pemrosesan yang mencakup penanganan nilai kosong dan tak hingga, *encoding* atribut kategorikal, serta normalisasi fitur. Data kemudian dibagi menjadi subset pelatihan dan pengujian sebelum diterapkan pada model Random Forest. Evaluasi model dilakukan menggunakan confusion matrix serta metrik akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa penerapan *Class Weight* mampu meningkatkan performa model secara signifikan, dengan capaian akurasi 99,98%, presisi 99,98%, *recall* 99,97%, dan F1-*score* 99,97%. Hasil ini membuktikan bahwa algoritma yang digunakan efektif dalam mendeteksi dan mengklasifikasikan serangan DDoS secara akurat.

Kata kunci: DDoS, random forest, class weight, CICDDoS2019

Classification of DDoS attacks using the random forest method and class weight technique on the CICDDoS2019 dataset

Abstract

The rapid advancement of information technology has significantly influenced various aspects of life, including an increasing reliance on network-based services. However, this dependence has also led to the emergence of more complex cybersecurity threats, one of the most prominent being Distributed Denial of Service (DDoS) attacks. These attacks can disrupt service availability by overwhelming target systems with excessive traffic. A major challenge in detecting DDoS attacks lies in the wide variety of attack patterns and the class imbalance that commonly occurs in network traffic datasets. To address these issues, a machine learning–based approach capable of handling complex attack behaviors while compensating for imbalanced data distribution is required. One potential solution is the use of the Random Forest algorithm with class-weight techniques, applied to the CICDDoS2019 dataset. The research procedure includes data collection and exploration, preprocessing steps such as handling missing and infinite values, encoding categorical attributes, and feature normalization. The dataset is then divided into training and testing subsets before being processed by the Random Forest model. Model evaluation is conducted using a confusion matrix along with accuracy, precision, recall, and F1-score metrics. Experimental results show that incorporating

class weight significantly improves model performance, achieving an accuracy of 99.98%, precision of 99.98%, recall of 99.97%, and an F1-score of 99.97%. These findings demonstrate that the proposed approach is highly effective for accurately detecting and classifying DDoS attacks.

Keywords: *DDoS, random forest, class weight, CICDDoS2019*

1. PENDAHULUAN

Teknologi informasi mendorong meningkatnya penggunaan internet di berbagai bidang, mulai dari pendidikan, pemerintahan, industri, hingga komunikasi. Namun, seiring dengan meningkatnya pemanfaatan teknologi informasi, risiko ancaman terhadap keamanan sistem informasi juga semakin besar. Salah satu ancaman yang paling umum dan berbahaya adalah *Distributed Denial of Service* (DDoS) *attack*, yaitu serangan yang membanjiri jaringan dengan lalu lintas berlebih sehingga sumber daya sistem tidak dapat diakses oleh pengguna yang sah [1]. Serangan ini biasanya melibatkan sejumlah sistem komputer yang secara bersamaan mengirimkan data dalam jumlah besar hingga server tidak mampu menangani permintaan yang datang [2].

Pada kuartal kedua tahun 2020, jumlah serangan DDoS tercatat meningkat dari 302,08% menjadi 316,67%, atau lebih dari tiga kali lipat dibandingkan periode yang sama pada tahun 2019 [3]. Kondisi ini menunjukkan urgensi upaya mitigasi yang lebih efektif, salah satunya melalui pemanfaatan *Intrusion Detection System* (IDS). IDS dirancang untuk memantau lalu lintas jaringan secara berkesinambungan dan mengidentifikasi aktivitas anomali yang berpotensi menjadi ancaman keamanan [4]. *Intrusion Detection System* (IDS) bekerja secara pasif, artinya sistem ini berfungsi mendeteksi adanya penyusup dan memberikan informasi kepada administrator jaringan apabila terjadi serangan atau gangguan [5]. dengan sasaran memonitoring aset jaringan sehingga dapat mengenali perilaku yang tidak lazim, kegiatan yang tidak sesuai, serangan atau upaya penyusupan, serta menyediakan informasi untuk menelusuri penyerang [6].

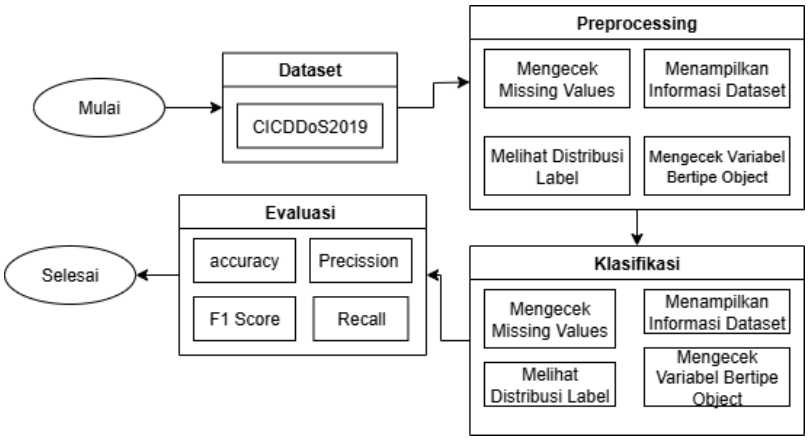
IDS berperan penting dalam menjaga aspek kerahasiaan, ketersediaan, dan integritas sistem [7]. Data yang dikumpulkan oleh IDS, seperti lalu lintas jaringan, catatan keamanan, dan aktivitas serangan, umumnya digunakan dalam penelitian *machine learning* untuk mengembangkan sistem deteksi yang lebih cerdas. Salah satu dataset IDS yang banyak dimanfaatkan adalah CICDDoS2019 [8].

Berbagai penelitian telah dilakukan menggunakan dataset ini [9] menggunakan algoritma XGBoost dengan teknik SMOTE untuk mengatasi masalah *imbalanced class* dan imputasi rata-rata untuk menangani missing value. Hasilnya, model mencapai akurasi 99% dengan nilai *precision, recall*, dan *F1-score* hingga 100%. Penelitian lain oleh [10]. menunjukkan bahwa penggunaan *class weight* pada dataset BotIoT mampu mencapai akurasi 100%. Selanjutnya, [11]. juga membuktikan bahwa penerapan Random Forest dengan *class weight, feature selection*, dan permutation importance menghasilkan *weighted F1-score* 99,8% dan macro *F1-score* 93,31% pada *dataset* CICIDS-2017.

Berdasarkan permasalahan tersebut, penelitian ini difokuskan pada klasifikasi serangan DDoS menggunakan algoritma Random Forest dengan penerapan *class weight* pada dataset CICDDoS2019. Tujuannya adalah untuk mengatasi ketidakseimbangan data sehingga model dapat menghasilkan klasifikasi yang lebih seimbang antara kelas mayoritas dan minoritas, serta meningkatkan akurasi deteksi khususnya pada serangan SYN Flood.

2. METODE PENELITIAN

Pada penelitian ini, terdapat beberapa tahap dalam proses klasifikasi serangan DDoS seperti pada Gambar 1.



Gambar 1. Kerangka Penelitian

Gambar 1 yaitu kerangka penelitian, menggambarkan struktur penelitian yang digunakan untuk melakukan klasifikasi serangan DDoS pada *dataset* CICDDoS2019. Tahapan penelitian dimulai dengan penginputan dataset, kemudian dilakukan proses

preprocessing yang meliputi pengecekan *missing values*, menampilkan informasi dataset, melihat distribusi label, serta mengecek variabel bertipe objek. Selanjutnya, dilakukan tahap klasifikasi menggunakan algoritma Random Forest dengan penerapan teknik *class weight* untuk menangani ketidakseimbangan kelas. Random Forest merupakan metode *ensemble* yang membentuk model prediksi melalui kombinasi sejumlah pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko *overfitting* [12][13]. Dan Teknik *class weight* digunakan dengan cara memberikan bobot yang lebih tinggi pada kelas minoritas [14]. Tahap akhir adalah evaluasi kinerja model menggunakan *confusion matrix* dengan metrik akurasi, *presisi, recall*, dan *F1-score* untuk menilai efektivitas metode klasifikasi yang diterapkan.

2.1. Dataset

Dataset yang digunakan dalam penelitian ini adalah CIC-DDoS2019, yang disediakan oleh Canadian Institute for Cybersecurity (CIC), University of New Brunswick, Kanada. Dataset ini tersedia secara publik [8]. Dataset ini dirancang untuk mendukung pengembangan sistem deteksi serangan Distributed Denial of Service (DDoS) dengan data yang diperoleh dari skenario lalu lintas jaringan realistis yang melibatkan aktivitas normal maupun aktivitas serangan.

Dalam penelitian ini, data yang digunakan difokuskan pada dua kelas, yaitu SYN (Serangan) dan Benign (Normal), dengan distribusi sebagai berikut:

Tabel 1. Distribusi Kelas Dataset CICDDoS2019

Kelas	Jumlah Datset
SYN (Serangan)	4.282.751
Benign (Normal)	35.790
Total	4.320.541

Tabel 1 yaitu Distribusi Kelas Dataset CIC-DDoS2019, di atas menunjukkan bahwa dataset memiliki masalah ketidakseimbangan kelas (imbalanced class). Jumlah data pada kelas SYN (Serangan) jauh lebih besar dibandingkan dengan kelas Benign (Normal), sehingga dapat memengaruhi kinerja algoritma klasifikasi apabila tidak ditangani dengan tepat.

Dataset CIC-DDoS2019 tersedia dalam format CSV, di mana setiap baris merepresentasikan satu flow jaringan yang terdiri atas 87 fitur. Fitur-fitur tersebut mencakup informasi statistik dari aliran data, seperti durasi, jumlah dan ukuran paket, kecepatan transfer, serta informasi terkait protokol, port, dan flag TCP/IP. Karakteristik ini memungkinkan dataset untuk digunakan dalam membedakan antara aktivitas normal (benign), benign merupakan jenis data lalu lintas jaringan yang menggambarkan aktivitas wajar, tidak berisiko, serta tidak termasuk dalam kategori serangan atau ancaman terhadap keamanan [15] dengan aktivitas berbahaya (serangan SYN).

Serangan SYN Flood sendiri merupakan salah satu bentuk serangan DDoS yang mengeksploitasi mekanisme three-way handshake pada protokol TCP. Penyerang mengirimkan sejumlah besar permintaan koneksi (SYN) tanpa pernah menyelesaikan proses handshake, sehingga sumber daya server menjadi terbebani dan akhirnya tidak mampu merespons permintaan koneksi yang sah. Hal inilah yang menjadikan serangan SYN Flood sangat berbahaya dan umum digunakan dalam skenario DDoS [16].

2.2. *Preprocessing*

Tahap *preprocessing* dilakukan untuk memastikan kualitas data sebelum masuk ke proses pelatihan model. *Preprocessing* merupakan tahapan awal yang dapat dilakukan untuk mempermudah dalam pengambilan maupun menerapkan pengambilan data, di mana data yang kurang sempurna maupun data yang tidak konsisten dapat disaring menjadi data yang diperlukan [17]. Langkah pertama adalah melakukan pemeriksaan awal terhadap *dataset* untuk memahami struktur data, ukuran, serta tipe variabel yang digunakan. Pemeriksaan ini juga mencakup identifikasi nilai kosong *(missing values)* guna memastikan tidak ada data yang hilang yang dapat memengaruhi hasil klasifikasi. Apabila ditemukan nilai kosong, penanganan dapat dilakukan dengan menghapus data yang bermasalah atau menggunakan teknik imputasi. Selain itu, juga dilakukan pengecekan terhadap nilai tak hingga *(infinite values),* baik positif maupun negatif, yang biasanya muncul akibat kesalahan perhitungan seperti pembagian dengan nol. Jika nilai tersebut ditemukan, maka perlu ditangani dengan menggantinya menggunakan nilai median atau mean pada kolom terkait, atau dengan menghapus baris yang bermasalah.

Selanjutnya, *preprocessing* digunakan untuk memeriksa distribusi label untuk mengetahui sebaran jumlah data pada setiap kelas. Hasil pengecekan menunjukkan bahwa jumlah data pada kelas SYN *Flood* mencapai 4.284.751 *instance,* sedangkan kelas *Benign* hanya berjumlah 35.790 *instance*. Distribusi ini memperlihatkan adanya ketidakseimbangan kelas yang signifikan, di mana kelas mayoritas (SYN *Flood*) jauh lebih mendominasi dibandingkan kelas *minoritas (Benign).* Kondisi ini berpotensi menimbulkan bias pada model klasifikasi, sehingga diperlukan teknik penanganan *imbalance* agar model mampu mengenali kedua kelas dengan lebih seimbang.

2.3. Klasifikasi

Pada tahap ini dilakukan proses klasifikasi untuk membedakan antara trafik normal *(Benign)* dan serangan SYN *Flood* pada *dataset* CICDDoS2019. Metode yang digunakan adalah algoritma *Random Forest* dengan penerapan *class weight* untuk

mengatasi masalah ketidakseimbangan kelas. *Random Forest* merupakan metode *ensemble learning* berbasis *bagging* yang menggabungkan sejumlah pohon keputusan *(decision tree)* untuk menghasilkan prediksi yang lebih akurat dan stabil. Selain itu, proses algoritma *Random Forest* bekerja dengan mengambil keputusan kelas terbanyak sesuai hasil dari pohon Keputusan [18]. Setiap pohon keputusan dalam *Random Forest* dilatih dengan sampel data yang diambil secara acak, dan hasil akhir klasifikasi ditentukan melalui mekanisme *voting* dari seluruh pohon yang terbentuk. Keunggulan *Random Forest* terletak pada kemampuannya menangani data berukuran besar, mengurangi risiko *overfitting,* serta mampu bekerja dengan baik meskipun terdapat fitur yang tidak relevan.

Dalam penelitian ini, salah satu tantangan utama adalah adanya *class imbalance*, di mana jumlah data pada kelas *mayoritas* (SYN *Flood*) jauh lebih besar dibandingkan kelas *minoritas (Benign)*[19]. Kondisi ini dapat menyebabkan model bias terhadap kelas *mayoritas* sehingga performa klasifikasi pada kelas *minoritas* menurun. Untuk mengatasi hal tersebut, diterapkan teknik *class weight,* yaitu memberikan bobot yang lebih besar pada kelas *minoritas* agar model lebih memperhatikan kelas tersebut saat proses pelatihan. Dengan demikian, prediksi menjadi lebih seimbang antara kedua kelas dan tidak hanya bergantung pada dominasi data *mayoritas.*

Secara keseluruhan, penerapan *Random Forest* dengan *class weight* diharapkan mampu meningkatkan kemampuan model dalam mendeteksi serangan SYN *Flood* sekaligus tetap mempertahankan akurasi yang baik pada trafik normal. Evaluasi hasil klasifikasi akan dilakukan menggunakan *confusion matrix* dan metrik performa lainnya seperti *presisi, recall,* dan *F1-score* untuk menilai sejauh mana model dapat mengenali kedua kelas secara seimbang.

2.4. Evaluasi Model

Evaluasi model *Random Forest* dengan *class weight* dalam penelitian ini dilakukan tidak hanya berdasarkan akurasi, tetapi juga metrik-metrik seperti *precision, recall*, dan *F1-score*. *Confusion matrix* digunakan untuk memperoleh nilai *True Positive, True Negative, False Positive,* dan *False Negative.* Pendekatan ini serupa dengan penelitian *Optimizing Performance Random Forest Algorithm sing Correlation-Based Feature Selection* yang menggunakan *dataset* CIC-DDoS2019 dan melaporkan peningkatan nilai akurasi bersama *precision, recall*, dan *F1-score* setelah seleksi fitur *statistic* [20].

3. Hasil dan Pembahasan

Eksperimen dilakukan untuk menguji kinerja model *Random Forest* dengan penerapan teknik *class weight* pada *dataset CICDDoS2019*. Tujuan dari eksperimen ini adalah untuk mengetahui bagaimana variasi pembagian data latih dan data uji *(train-test split)* memengaruhi performa model, khususnya dalam mendeteksi serangan DDoS jenis SYN. *Dataset* dibagi menggunakan metode *stratified split* agar distribusi kelas tetap proporsional pada data latih maupun data uji. Evaluasi performa dilakukan menggunakan *metrik akurasi, precision, recall,* dan *F1-score*.

Tabel 2 berikut menampilkan hasil evaluasi kinerja algoritma *Random Forest* dengan penerapan *class weight* pada berbagai skenario pembagian data latih dan data uji, menggunakan metrik *Accuracy, Precision, Recall,* dan *F1-Score.*

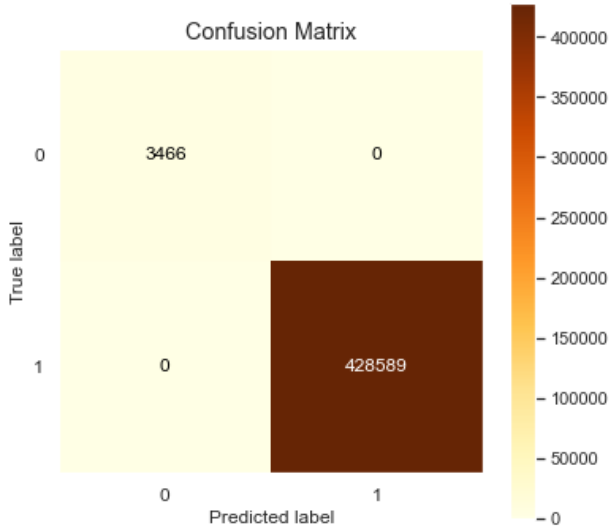
Tabel 2 Eksperimen Model Random Frest dengan Teknik Class Weight

Data Latih	Data Uji	*Accuracy*	*Precision*	*Recall*	*F1-Score*
90	10	100.00000%	1.00000	1.00000	1.00000
			1.00000	1.00000	1.00000
80	20	99.99965%	0.99958	1.00000	0.99979
			1.00000	1.00000	1.00000
70	30	99.99969%	0.99972	0.99991	0.99981
			1.00000	1.00000	1.00000
60	40	99.99925%	0.99916	0.99993	0.99954
			1.00000	0.99999	1.00000

Berdasarkan tabel di atas, terlihat bahwa keempat skenario menghasilkan performa yang sangat baik dengan nilai akurasi di atas 99% pada setiap percobaan. Namun, skenario terbaik diperoleh pada pembagian data 90% latih dan 10% uji, yang menghasilkan akurasi sempurna sebesar 100% dan metrik evaluasi lainnya juga mencapai nilai 1.00000. Hal ini menunjukkan bahwa model mampu belajar secara maksimal dengan jumlah data latih yang lebih besar, sekaligus tetap mempertahankan kemampuan generalisasi pada data uji.

Untuk memperkuat hasil, evaluasi dilakukan menggunakan *confusion matrix* yang memberikan gambaran lebih detail mengenai distribusi prediksi model terhadap kelas *BENIGN* (0) dan SYN (1). *Confusion matrix* mampu menunjukkan jumlah prediksi yang

benar maupun salah, serta membantu mengidentifikasi kemungkinan terjadinya kesalahan berupa *false positive* dan *false negative*.



Gambar 2 Hasil Confusion Matrix

Berdasarkan Gambar 2, yaitu Hasil *Confusion Matrix* dapat dilihat bahwa model berhasil mengklasifikasikan seluruh data uji dengan sempurna. Sebanyak 3.466 data benign diprediksi benar sebagai benign *(True Negative),* dan 428.589 data SYN diprediksi benar sebagai SYN *(True Positive).* Tidak terdapat kesalahan prediksi pada kedua kelas, sehingga tidak ditemukan *false positive* maupun *false negative*. Hal ini tercermin pada metrik evaluasi berupa akurasi, *precision, recall,* dan *F1-score* yang semuanya mencapai nilai maksimal, yakni 1.00000.
Hasil ini membuktikan bahwa penggunaan algoritma *Random Forest* dengan teknik *class weight* sangat efektif dalam menangani ketidakseimbangan kelas pada dataset CICDDoS2019. Dengan bobot tambahan pada kelas *minoritas,* model mampu belajar secara seimbang sehingga tidak hanya fokus pada kelas *mayoritas,* tetapi juga dapat mengenali kelas *minoritas* dengan tingkat akurasi yang sama tinggi. Dengan demikian, dapat disimpulkan bahwa metode ini memiliki potensi besar untuk diterapkan dalam sistem deteksi dini serangan DDoS yang membutuhkan performa tinggi serta keandalan dalam menangani distribusi data yang tidak seimbang.

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa algoritma Random Forest dengan penerapan class weight terbukti sangat efektif dalam menangani permasalahan ketidakseimbangan kelas pada dataset CICDDoS2019. Hasil eksperimen pada berbagai skenario pembagian data (90:10, 80:20, 70:30, dan 60:40) menunjukkan performa model yang konsisten tinggi, dengan akurasi selalu berada di atas 99% serta nilai *precision, recall*, dan *F1-score* yang mendekati 1,00000. Skenario terbaik diperoleh pada konfigurasi 90% data latih dan 10% data uji, di mana model mencapai akurasi sempurna 100% dengan seluruh metrik evaluasi lainnya juga berada pada nilai maksimum. Analisis menggunakan *confusion matrix* menegaskan tidak adanya kesalahan prediksi, baik *false positive* maupun *false negative*. Hal ini menunjukkan bahwa model mampu mengenali kelas mayoritas (SYN) maupun kelas minoritas (*Benign*) secara seimbang dan akurat. Temuan ini membuktikan bahwa kombinasi Random Forest dan teknik *class weight* merupakan pendekatan yang andal dan sangat efektif untuk mendeteksi serangan DDoS, khususnya jenis SYN, pada dataset dengan tingkat ketidakseimbangan yang sangat tinggi.

DAFTAR PUSTAKA

[1] Wahyuni and Pitrasacha Adytia, "Perbandingan Algoritma Machine Learning Dalam Mendeteksi Serangan DDOS," *Tematik*, vol. 9, no. 2, pp. 161–166, 2022, doi: 10.38204/tematik.v9i2.1070.

[2] Y. I. Mahendra and R. E. Putra, "Penerapan Algoritma Gradient Boosted Decision Tree (GBDT) untuk Klasifikasi Serangan DDoS," *JINACS (Journal Informatics Comput. Sci. ISSN*, vol. 06, pp. 158–166, 2024.

[3] J. A. Perez-diaz and J. A. Cantoral-ceballos, "Transport and Application Layer DDoS Attacks Detection to IoT," 2022.

[4] N. Dat-thinh, H. Xuan-ninh, and L. Kim-hung, "MidSiot : A Multistage Intrusion Detection System for," vol. 2022, no. December 2017, 2022, doi: 10.1155/2022/9173291.

[5] I. Riadi, D. Mualfah, and I. Riadi, "Network Forensics for Detecting Flooding Attack on Web Server," *Int. J. Comput. Sci. Inf. Secur.*, vol. 15, no. 2, pp. 326–331, 2017.

[6] M. Muqorobin, Z. Hisyam, M. Mashuri, H. Hanafi, and Y. Setiyantara, "Implementasi Network Intrusion Detection System (NIDS) Dalam Sistem Keamanan Open Cloud Computing," *Maj. Ilm. Bahari Jogja*, vol. 17, no. 2, pp. 1–9, 2019, doi: 10.33489/mibj.v17i2.205.

[7] M. Aljanabi, M. A. Ismail, and A. H. Ali, "Intrusion detection systems, issues, challenges, and needs," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 560–571, 2021, doi: 10.2991/ijcis.d.210105.001.

[8] I. Sharafaldin, A. H. Lashkari, and S. H. and A. A. G. (isharafa;, "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy," *ACM Int. Conf. Proceeding Ser.*, no. Cic, pp. 70–75, 2019, doi: 10.1145/3340997.3341005.

[9] R. Gunawan, E. S. Handika, and E. Ismanto, "Pendekatan Machine Learning Dengan Menggunakan Algoritma Xgboost (Extreme Gradient Boosting) Untuk Peningkatan Kinerja Klasifikasi Serangan Syn," vol. 3, no. 2, pp. 358–366, 2022.

[10] J. Al Amien, H. A. Ghani, N. I. M. Saleh, E. Ismanto, and R. Gunawan, "Intrusion detection system for imbalance ratio class u sing weighted XGBoost classifier," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 21, no. 5, pp. 1102–1112, 2023, doi: 10.12928/TELKOMNIKA.v21i5.24735.

[11] M. T. Abdelaziz *et al.*, *Enhancing Network Threat Detection with Random Forest-Based NIDS and Permutation Feature Importance*, vol. 33, no. 1. Springer US, 2025. doi: 10.1007/s10922-024-09874-0.

[12] D. Mualfah, W. Fadila, and R. Firdaus, "Teknik SMOTE untuk Mengatasi Imbalance Data pada Deteksi Penyakit Stroke Menggunakan Algoritma Random Forest," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 2, pp. 107–113, 2022, doi: 10.37859/coscitech.v3i2.3912.

[13] D. Mualfah, A. Prihatin, R. Firdaus, and Sunanto, "Analisis Sentimen Masyarakat Terhadap Kasus Pembobolan Data Nasabah Bank BSI Pada Twitter Menggunakan Metode Random Forest Dan Naïve Bayes," *J. Fasilkom*, vol. 13, no. 3, pp. 614–620, 2024, doi: 10.37859/jf.v13i3.6478.

[14] B. BAKIRARAR and S. YILMAZ IŞIKHAN, "A New Class-Weighting Formulation for the Class Imbalance Problem: A Methodological Research," *Turkiye Klin. J. Biostat.*, vol. 15, no. 2, pp. 79–90, 2023, doi: 10.5336/biostatic.2023-96293.

[15] M. Andrecut, "Attack vs Benign Network Intrusion Traffic Classification," no. 2, pp. 1–8, 2022, [Online]. Available: http://arxiv.org/abs/2205.07323

[16] D. Scholz, S. Gallenmüller, H. Stubbe, B. Jaber, M. Rouhi, and G. Carle, "Me love (SYN-)cookies: SYN flood mitigation in programmable data planes," *arXiv*, 2020.

[17] S. Nanda, D. Mualfah, and D. A. Fitri, "Analisis Sentimen Kepuasan Pengguna Terhadap Layanan Streaming Mola Menggunakan Algoritma Random Forest," *J. Apl. Teknol. Inf. dan Manaj.*, vol. 3, no. 2, pp. 210–219, 2022, doi: 10.31102/jatim.v3i2.1592.

[18] F. T. Admojo, S. Risnanto, A. W. Windiawati, M. Innuddin, and D. Mualfah, "Comparison of Naïve Bayes and Random Forest Algorithm in Webtoon Application Sentiment Analysis," *Innov. Res. Informatics*, vol. 6, no. 1, pp. 23–28, 2024, doi: 10.37058/innovatics.v6i1.10636.

[19] P. Y. Saputra, M. Z. Abdullah, and A. P. Kirana, "Improvisasi Teknik Oversampling MWMOTE Untuk Penanganan Data Tidak Seimbang," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 398, 2021, doi: 10.30865/mib.v5i2.2811.

[20] S. Soim, S. Sholihin, and C. B. Subianto, "Optimizing Performance Random Forest Algorithm Using Correlation-Based Feature Selection (CFS) Method to Improve Distributed Denial of Service (DDoS) Attack Detection Accuracy," *Indones. J. Artif. Intell. Data Min.*, vol. 7, no. 2, p. 220, 2024, doi: 10.24014/ijaidm.v7i2.24783.