Scan
Jurnal Teknologi Informasi dan Komunikasi

ISSN (Print): 1978-0087
ISSN (Online): 2686-6099
Volume 20 Nomor 1, Februari 2025
Halaman : 12 - 21

Perbandingan Kinerja CD-KNN Dalam Klasifikasi Serangan *Ddos UDP Flood*

Muhammad Bagus Satrio, Fetty Tri Anggraeny*, Achmad Junaidi
Program Studi Infomatika, Fakultas Ilmu Komputer,
Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia

Diterima: 18 Desember 2024 | Revisi: 2 Januari 2025 | Diterbitkan: 1 Februari 2025

DOI: https://doi.org/10.33005/scan.v20i1.5387

ABTSRAK

Keamanan siber menjadi isu kritis di era digital, terutama serangan Distributed Denial of Service (DDoS) UDP Flood yang meningkat signifikan hingga 59,31% pada tahun 2023. Penelitian ini mengimplementasikan algoritma Cluster-based Dynamic K-Nearest Neighbor (CD-KNN) dengan parameter α = 6, yang diperkuat dengan seleksi fitur Analysis of Variance (ANOVA) dan teknik penyeimbangan data Synthetic Minority Oversampling Technique (SMOTE) untuk mendeteksi serangan DDoS UDP Flood berbasis anomali lalu lintas jaringan. Dataset CICDDoS2019 diproses melalui tahapan preprocessing menyeluruh, termasuk penghapusan 20 fitur tidak relevan, normalisasi Min-Max, dan penyeimbangan distribusi kelas. Evaluasi dilakukan pada dua skenario: CD-KNN dengan ANOVA dan CD-KNN dengan ANOVA-SMOTE. Hasil eksperimen menunjukkan bahwa model CD-KNN dengan ANOVA-SMOTE menghasilkan performa terbaik dengan akurasi 99.994%, presisi 99.998%, recall 99.996%, dan F1-score 99.997%. Penerapan SMOTE meningkatkan recall kelas minoritas sebesar 2,47% dibandingkan model tanpa SMOTE, dengan waktu komputasi 3903 detik yang tetap efisien untuk dataset berskala besar. Temuan ini menunjukkan bahwa kombinasi CD-KNN, ANOVA, dan SMOTE merupakan pendekatan komprehensif yang efektif dalam sistem deteksi intrusi berbasis anomali untuk serangan DDoS UDP Flood.

Kata Kunci: *DDoS UDP Flood, CD-KNN, ANOVA, SMOTE*, Deteksi Anomali.

Comparison Of CD-KNN Performance With And Without ANOVA-SMOTE In Classifying UDP Flood Ddos Attacks

ABSTRACT

Cybersecurity has become a critical issue in the digital era, particularly with the significant rise in Distributed Denial of Service (DDoS) UDP Flood attacks, reaching 59.31% in 2023. This study implements the Cluster-based Dynamic K-Nearest Neighbor (CD-KNN) algorithm with α = 6, enhanced with Analysis of Variance (ANOVA) feature selection and the Synthetic Minority Oversampling Technique (SMOTE), for anomaly-based detection of DDoS UDP Flood attacks. The CICDDoS2019 dataset underwent comprehensive preprocessing, including the removal of 20 irrelevant features, Min-Max normalization, and class distribution balancing. Performance was evaluated in two scenarios: CD-KNN with ANOVA and CD-KNN with ANOVA-SMOTE. Experimental results show that CD-KNN with ANOVA-SMOTE achieved the best performance, reaching 99.994% accuracy, 99.998% precision, 99.996% recall, and 99.997% F1-score. The use of SMOTE improved minority class recall by 2.47% compared to the model without SMOTE, with a computation time of 3903 seconds, which is efficient for large-scale datasets. These findings demonstrate the effectiveness of combining CD-KNN, ANOVA, and SMOTE as a comprehensive solution for anomaly-based intrusion detection systems against DDoS UDP Flood attacks.

Keywords: *DDoS UDP Flood, CD-KNN, ANOVA, SMOTE, Anomaly Detection*

***Corresponding Author:**
Email : fettyanggraeny.if@upnjatim.ac.id
Alamat : Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294

PENDAHULUAN

Keamanan siber menjadi isu strategis di era digital modern, terutama dengan meningkatnya frekuensi dan kompleksitas serangan siber seperti Distributed Denial of Service (DDoS). Salah satu bentuk serangan DDoS yang paling banyak terjadi adalah serangan UDP Flood, yang membanjiri sistem target dengan paket UDP hingga menyebabkan gangguan layanan. Menurut laporan Qrator Labs, metode UDP Flood menyumbang hingga 59,31% dari total serangan DDoS pada kuartal kedua tahun 2023 (Qrator Labs, 2024). Di Indonesia, laporan dari SOCRadar mencatat lebih dari 43.000 serangan DDoS selama tahun 2024, dengan puncak intensitas mencapai 693 Gbps, menjadikan metode ini sebagai ancaman utama bagi ketersediaan layanan digital nasional (SOC Radar, 2024).

Deteksi dini terhadap serangan sangat penting, dan pendekatan berbasis anomali menjadi metode yang banyak digunakan karena mampu mengenali pola lalu lintas yang menyimpang dari normal. Salah satu algoritma yang handal dalam klasifikasi berbasis anomali adalah Cluster-based Dynamic K-Nearest Neighbor (CD-KNN). Algoritma ini merupakan pengembangan dari K-Nearest Neighbor (KNN) klasik, dengan keunggulan utama pada penentuan nilai K yang bersifat dinamis, disesuaikan berdasarkan kepadatan lokal tiap klaster (Robindro et al., 2022). Dengan demikian, CD-KNN dapat mengatasi permasalahan pemilihan nilai K statis yang dapat menyebabkan bias klasifikasi, serta lebih adaptif terhadap distribusi data yang kompleks dan tidak seimbang Selain pemilihan algoritma, kualitas data juga berperan penting dalam performa model klasifikasi. Salah satu teknik untuk meningkatkan kualitas data adalah dengan melakukan seleksi fitur menggunakan Analysis of Variance (ANOVA), yang mampu mengidentifikasi fitur yang paling relevan dengan target klasifikasi (Chandrashekar & Sahin, 2014). Untuk mengatasi ketimpangan kelas dalam data di mana jumlah data serangan jauh lebih sedikit dibandingkan data normal digunakan teknik Synthetic Minority Oversampling Technique (SMOTE). Teknik ini menghasilkan sampel sintetis untuk kelas minoritas guna menyeimbangkan distribusi kelas dan meningkatkan sensitivitas model terhadap deteksi serangan (Chawla et al., 2002).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengimplementasikan dan membandingkan kinerja algoritma CD-KNN dalam dua skenario: dengan seleksi fitur ANOVA dan tanpa penyeimbangan data, serta dengan kombinasi ANOVA-SMOTE. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score, untuk menilai efektivitas pendekatan ini dalam mendeteksi serangan DDoS UDP Flood berbasis anomali lalu lintas jaringan.

METODE PENELITIAN

Bagian ini menjelaskan secara detail setiap tahap yang dilalui selama proses penelitian. Pada Gambar 1 menunjukkan alur metode penelitian yang diusulkan dalam studi ini. Proses dimulai dari tahap pengumpulan data, di mana dataset serangan DDoS UDP Flood berbasis anomali lalu lintas jaringan diperoleh dari sumber tepercaya. Data yang terkumpul selanjutnya melalui tahap preprocessing untuk menghilangkan nilai hilang, duplikasi, dan inkonsistensi, sehingga memastikan kualitas data yang optimal sebelum proses analisis. Tahap berikutnya adalah seleksi fitur, yang bertujuan untuk memilih subset atribut paling relevan melalui metode statistik sehingga kompleksitas model dapat dikurangi dan kinerja klasifikasi ditingkatkan. Data yang telah diseleksi kemudian

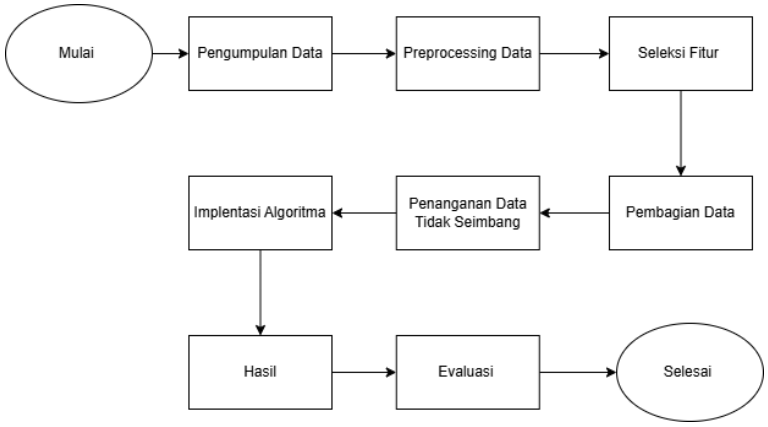
melalui pembagian data menjadi data pelatihan dan data pengujian dengan proporsi tertentu. Mengingat dataset memiliki distribusi kelas yang tidak seimbang, dilakukan penanganan data tidak seimbang menggunakan metode oversampling berbasis Synthetic Minority Over-sampling Technique (SMOTE) untuk meningkatkan representasi kelas minoritas. Selanjutnya, tahap implementasi algoritma dilakukan dengan menerapkan Cluster-based Dynamic K-Nearest Neighbor (CD-KNN) sebagai metode klasifikasi utama. Model yang dihasilkan kemudian dievaluasi pada tahap evaluasi menggunakan metrik kinerja seperti akurasi, presisi, recall, dan F1-score. Proses ini menghasilkan temuan yang menjadi dasar penarikan kesimpulan terkait efektivitas pendekatan yang diusulkan terhadap deteksi serangan DDoS UDP Flood.

Pengumpulan Data

Penelitian ini menggunakan data dari dataset CICDDoS 2019 (Canadian Institute for Cybersecurity Distributed Denial of Service 2019) untuk menganalisis serangan DDoS jenis UDP Flood. Dataset ini dipilih karena mencakup beragam tipe serangan DDoS, termasuk UDP Flood, yang dikumpulkan dalam lingkungan yang realistis serta disertai pelabelan yang akurat (Sharafaldin et al., 2019). Selain itu, CICDDoS2019 juga memuat data normal (benign) dan serangan DDoS yang umum terjadi, yang dihasilkan dari lalu lintas jaringan nyata dengan mencakup berbagai variasi serangan berbasis protokol TCP maupun UDP (Salmi & Oughdir, 2023).

Preprocessing Data

Pre-processing merupakan tahapan penting dalam analisis data yang bertujuan untuk meningkatkan kualitas data guna efisiensi dan akurasi proses data mining (Alassaf & Qamar, 2022) . Proses ini dimulai dengan penghapusan fitur yang tidak relevan, seperti atribut string atau nilai konstan, untuk mengurangi dimensi data. Selanjutnya, dilakukan pembersihan data yang mencakup penanganan nilai yang hilang (missing values), penghapusan nilai negatif, serta duplikasi data. Untuk data klasifikasi, label kategoris seperti 'Benign' dan 'DrDoS_UDP' dikonversi menjadi format numerik (misalnya, 0 dan 1) melalui label encoding. Terakhir, data dinormalisasi menggunakan Min-Max Scaling untuk menyamakan rentang nilai fitur numerik ke skala [0, 1], memastikan keseragaman data sebelum dilakukan pemodelan. Seluruh tahapan ini krusial untuk menghasilkan model yang andal dan akurat.



Gambar 1. Metode Penelitian

Seleksi Fitur ANOVA

Analysis of Variance (ANOVA) adalah metode seleksi fitur yang digunakan untuk mengidentifikasi fitur-fitur paling relevan dalam membangun model klasifikasi. Dalam pendekatan ini, setiap fitur dievaluasi menggunakan nilai F-Score, yang mencerminkan seberapa besar perbedaan antar kelompok data dibandingkan dengan variasi di dalam masing-masing kelompok (Kumar et al., 2015). Tujuan dari ANOVA adalah untuk menentukan apakah terdapat perbedaan yang signifikan secara statistik antara rata-rata dari tiga atau lebih kelompok yang bersifat independen (tidak saling berhubungan) (Musthafa et al., 2024). Jumlah atribut yang berlebihan berpotensi memicu peningkatan kesalahan deteksi akibat adanya redundansi atau duplikasi data. Dengan menyederhanakan jumlah atribut, kinerja sistem dapat ditingkatkan (Megantara & Ahmad, 2020).

Pembagian Data

Pembagian dataset merupakan prosedur yang dianggap sangat penting untuk menghilangkan atau meminimalkan bias pada data pelatihan dalam model pembelajaran mesin (Muraina, 2022). Setelah tahap seleksi fitur selesai dilaksanakan, dataset dibagi menjadi dua subset, yaitu data pelatihan (training data) dan data pengujian (testing data). Pembagian ini bertujuan untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data yang belum pernah diproses sebelumnya (Varoquaux & Colliot, 2023). Pada penelitian ini, pembagian dilakukan dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian. Rasio 80:20 ini dipilih karena direkomendasikan pada dataset berukuran besar, sehingga jumlah data latih yang tersedia lebih memadai untuk menghasilkan performa klasifikasi yang lebih baik (Rácz et al., 2021).

SMOTE (Synthetic Minority Oversampling Technique)

SMOTE (Synthetic Minority Oversampling Technique) merupakan metode oversampling yang dirancang untuk menangani permasalahan ketidakseimbangan kelas dalam dataset. Ketidakseimbangan kelas terjadi ketika distribusi data antar kelas tidak merata, dengan kelas minoritas yang memiliki jauh lebih sedikit sampel dibandingkan kelas mayoritas (Elreedy et al., 2024). Algoritma SMOTE menghasilkan sampel sintetis dari kelas minoritas dengan melakukan interpolasi linier secara acak antara beberapa sampel dan tetangga terdekatnya (Wang et al., 2021). Dengan demikian, distribusi kelas menjadi lebih seimbang sehingga model pembelajaran tidak condong terhadap kelas mayoritas. Selain itu, penerapan SMOTE dapat meningkatkan kinerja prediksi pada dataset yang tidak seimbang (Joloudari et al., 2023).

CD-KNN

CD-KNN merupakan pengembangan dari algoritma K-Nearest Neighbor yang menggabungkan konsep clustering berdasrakan label dan penyesuaian nilai k secara dinamis berdasarkan kepadatan data. Pendekatan ini dirancang untuk meningkatkan efisiensi dan akurasi dalam klasifikasi, terutama dalam konteks deteksi intrusi jaringan. CD-KNN memanfaatkan teknik klasterisasi untuk membagi data latih ke dalam beberapa cluster, sehingga pencarian tetangga terdekat dilakukan secara terbatas hanya dalam cluster yang relevan. Selain itu, nilai k ditentukan secara adaptif sesuai dengan karakteristik lokal data uji, seperti jarak dan kepadatan (Robindro et al., 2022).

Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan confusion matrix, yang menyajikan informasi dalam bentuk tabel mengenai jumlah prediksi yang benar maupun salah untuk setiap kelas, termasuk True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) (Markoulidakis et al., 2021). Berdasarkan nilai-nilai ini, berbagai metrik evaluasi digunakan untuk mengukur akurasi klasifikasi, seperti presisi, sensitivitas (sensitivity), dan spesifisitas (specificity) (Ramli et al., 2022). Metrik lainnya seperti akurasi, recall, F1-score, dan ROC-AUC. Akurasi mengukur seberapa besar proporsi prediksi yang tepat dibandingkan dengan seluruh prediksi yang dilakukan. Presisi menggambarkan tingkat relevansi dari hasil prediksi positif. Recall menunjukkan sejauh mana model mampu mengenali seluruh data positif yang sebenarnya. F1-score merupakan rata-rata harmonis antara presisi dan recall, yang memberikan keseimbangan antara keduanya. Evaluasi model merupakan tahap penting untuk memastikan bahwa model yang dilatih tidak hanya menunjukkan performa baik pada data pelatihan, tetapi juga mampu melakukan generalisasi secara efektif terhadap data baru yang belum pernah digunakan sebelumnya (Heydarian et al., 2022)

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengimplementasikan algoritma Cluster-based Dynamic K-Nearest Neighbor (CD-KNN) dalam klasifikasi serangan Distributed Denial of Service (DDoS) jenis UDP Flood berbasis deteksi anomali. Fokus utama dari penelitian ini adalah mengintegrasikan metode seleksi fitur Analysis of Variance (ANOVA) guna meningkatkan efektivitas dalam pemilihan fitur yang relevan, serta menggunakan Synthetic Minority Over-sampling Technique (SMOTE) sebagai metode penyeimbang data untuk menangani ketidakseimbangan kelas dalam dataset. Model klasifikasi yang dikembangkan akan dievaluasi menggunakan berbagai metrik performa, seperti akurasi, presisi, recall, dan F1-score. Sebelum diterapkan dalam proses pelatihan dan pengujian, dataset telah melalui tahapan prapemrosesan, yang mencakup pembersihan data, normalisasi, dan encoding label. Hasil evaluasi disajikan dalam bentuk tabel dan grafik guna menunjukkan kinerja CD-KNN dalam mendeteksi serangan DDoS berdasarkan kombinasi fitur yang telah terseleksi serta data yang telah diseimbangkan. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem deteksi dini terhadap ancaman siber, khususnya pada lalu lintas jaringan yang menunjukkan pola anomali. Dengan menggabungkan seleksi fitur dan teknik balancing data, model yang diusulkan memiliki potensi untuk meningkatkan akurasi deteksi dibandingkan dengan pendekatan konvensional.

Hasil Pengujian

Setelah algoritma Cluster-based Dynamic K-Nearest Neighbor (CD-KNN) diimplementasikan dengan dukungan metode seleksi fitur ANOVA serta teknik penyeimbangan data SMOTE, dilakukan proses pengujian untuk menilai kinerja model dalam mengklasifikasikan serangan Distributed Denial of Service (DDoS) jenis UDP Flood. Evaluasi dilakukan dengan membandingkan hasil klasifikasi model terhadap label asli data, menggunakan sejumlah metrik performa seperti akurasi, presisi, recall, F1-score, dan ROC-AUC. Penggunaan kelima metrik ini memberikan pandangan yang komprehensif terhadap efektivitas model; di mana akurasi yang tinggi mencerminkan

ketepatan prediksi secara keseluruhan, sedangkan nilai presisi dan recall yang tinggi menunjukkan kemampuan model dalam mengenali serangan secara akurat dan menyeluruh. F1-score, yang merupakan rata-rata harmonis dari presisi dan recall, memberikan indikator keseimbangan kinerja model, khususnya dalam menangani data dengan distribusi kelas yang tidak merata. Hasil evaluasi dari berbagai skenario pengujian, baik sebelum maupun sesudah penerapan teknik balancing data SMOTE, disajikan melalui tabel dan grafik untuk menggambarkan performa klasifikasi dalam berbagai kombinasi fitur dan kondisi dataset. Tabel-tabel ini menjadi acuan dalam analisis lanjutan mengenai dampak seleksi fitur dan penyeimbangan data terhadap akurasi dan efisiensi model dalam mendeteksi serangan DDoS. Tabel 1 menyajikan hasil evaluasi model pada kondisi tanpa penerapan metode balancing data SMOTE.

Disimpulkan bahwa algoritma CD-KNN mampu memberikan performa klasifikasi yang sangat tinggi dalam mendeteksi serangan DDoS jenis UDP Flood, meskipun tanpa penerapan teknik penyeimbangan data. Secara umum, seluruh metrik evaluasi seperti akurasi, presisi, recall, dan F1-score berada di atas 99.98%, menunjukkan bahwa model dapat mengidentifikasi serangan dengan sangat baik. Di antara seluruh skenario yang diuji, nilai parameter α = 8 menunjukkan performa yang paling optimal. Pada konfigurasi ini, model mencatatkan akurasi sebesar 99.993%, presisi dan recall sebesar 99.994%, serta F1-score sebesar 99.996%. Selain itu, nilai ROC-AUC yang dicapai adalah 99.08, menunjukkan kemampuan tinggi dalam membedakan antara trafik serangan dan trafik normal. Keunggulan lain dari parameter ini adalah waktu komputasi yang lebih cepat, yaitu 2055 detik, menjadikannya lebih efisien dibandingkan parameter lain dengan performa serupa.

Sebaliknya, pada parameter a = 10, meskipun metrik akurasi dan F1-score tetap tinggi, terjadi penurunan pada nilai ROC-AUC menjadi 98.17, yang mengindikasikan penurunan kemampuan generalisasi model terhadap data baru. Waktu komputasi juga lebih rendah, tetapi dengan pengorbanan performa. Dengan demikian, pemilihan parameter a yang tepat sangat mempengaruhi keseimbangan antara akurasi klasifikasi dan efisiensi komputasi. Hasil ini akan menjadi referensi penting untuk perbandingan performa model setelah penerapan teknik SMOTE, yang disajikan pada Tabel 2 dalam penelitian ini.

Tabel 1.
Evaluasi Tanpa *SMOTE*

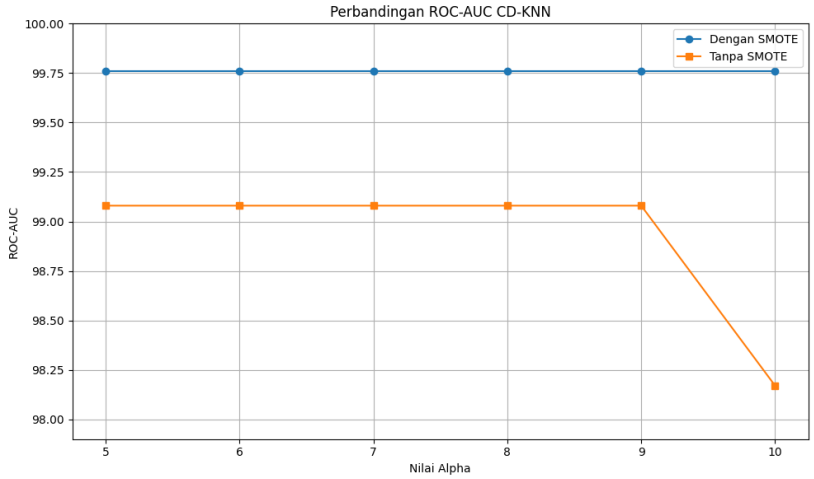
Parameter	Akurasi	Presisi	*Recall*	*F1-Score*	Nilai ROC-AUC	Waktu Komputasi (detik)
α = 5	99.992	99.994	99.997	99.996	99.08	2072 Detik
α= 6	99.992	99.994	99.997	99.996	99.08	2083 Detik
α = 7	99.992	99.994	99.997	99.996	99.08	2067 Detik
α = 8	99.993	99.994	99.998	99.996	99.08	2055 Detik
α = 9	99.993	99.994	99.998	99.996	99.08	2066 Detik
α = 10	99.988	99.989	99.998	99.994	98.17	2043 Detik

Sumber : Data Diolah

Tabel 2.
Evaluasi dengan *SMOTE*

Parameter	Akurasi	Presisi	*Recall*	*F1-Score*	Nilai ROC-AUC	Waktu Komputasi (detik)
α = 5	99.994	99.998	99.996	99.997	99.76	4016 Detik
α = 6	99.994	99.998	99.996	99.997	99.76	3903 Detik
α = 7	99.988	99.998	99.989	99.994	99.76	3984 Detik
α = 8	99.988	99.998	99.989	99.994	99.76	4001 Detik
α = 9	99.988	99.998	99.989	99.994	99.76	3991 Detik
α = 10	99.988	99.998	99.989	99.994	99.76	3996 Detik

Sumber : Data Diolah



Gambar 2. Grafik Perbandingan ROC-AUC

Setelah melakukan pengujian model CD-KNN tanpa balancing data, tahap selanjutnya adalah pengujian model dengan penerapan teknik SMOTE (Synthetic Minority Oversampling Technique) untuk mengatasi ketidakseimbangan kelas pada dataset. Rancangan pengujian ini dilakukan dengan tetap mempertahankan proses preprocessing, seleksi fitur menggunakan ANOVA, serta konfigurasi parameter a yang bervariasi, yakni dari α = 5 hingga α = 10. Penerapan SMOTE bertujuan untuk meningkatkan representasi kelas minoritas dalam data pelatihan, sehingga model dapat lebih seimbang dalam mengenali pola dari kedua kelas, baik serangan maupun non-serangan.

Evaluasi dilakukan menggunakan metrik yang sama seperti pada pengujian sebelumnya, yaitu akurasi, presisi, recall, F1-score, ROC-AUC, serta waktu komputasi. Berdasarkan hasil pada Tabel 2, terlihat bahwa parameter α = 6 menghasilkan performa terbaik dengan akurasi, presisi, recall, dan F1-score sebesar 99.994%–99.997%, serta nilai ROC-AUC tertinggi yaitu 99.76, disertai waktu komputasi tercepat sebesar 3903 detik. Hal ini menunjukkan bahwa penerapan SMOTE berhasil meningkatkan generalisasi model dan stabilitas prediksi, terutama terhadap kelas minoritas. Jika dibandingkan dengan hasil pengujian tanpa SMOTE, performa model CD-KNN secara

umum tetap tinggi, namun menjadi lebih konsisten pada semua metrik evaluasi, khususnya pada nilai ROC-AUC yang menunjukkan peningkatan.

Pada Gambar 2 menunjukkan grafik perbandingan nilai ROC-AUC dari algoritma CD-KNN (Cluster-based Dynamic K-Nearest Neighbor) dengan dan tanpa penerapan metode SMOTE pada berbagai nilai parameter alpha. Parameter alpha dalam CD-KNN berperan dalam menentukan batas kedekatan antara data uji dan pusat klaster, sehingga mempengaruhi pemilihan tetangga terdekat yang relevan dalam proses klasifikasi.

Berdasarkan grafik, terlihat bahwa model CD-KNN dengan SMOTE menghasilkan nilai ROC-AUC yang sangat tinggi dan stabil di sekitar 99,76% pada seluruh variasi alpha (5 hingga 10). Hal ini menunjukkan bahwa penerapan SMOTE secara konsisten menjaga kualitas klasifikasi CD-KNN, dengan performa yang tidak terpengaruh oleh perubahan nilai alpha.

Sementara itu, model CD-KNN tanpa SMOTE memperlihatkan performa yang sedikit lebih rendah dan mengalami penurunan drastis pada alpha = 10. Nilai ROC-AUC bertahan di kisaran 99,07% dari alpha 5 hingga 9, namun menurun menjadi sekitar 98,37% pada alpha 10. Penurunan ini mengindikasikan bahwa tanpa SMOTE, model menjadi lebih sensitif terhadap peningkatan nilai alpha, yang kemungkinan menyebabkan pemilihan tetangga yang kurang representatif akibat ketidakseimbangan kelas.

SIMPULAN

Penelitian ini menjelaskan bahwa algoritma CD-KNN mampu memberikan performa yang baik dalam mendeteksi serangan DDoS UDP Flood berbasis anomali lalu lintas jaringan. Hasil pengujian menunjukkan bahwa penerapan seleksi fitur menggunakan ANOVA serta teknik oversampling SMOTE secara signifikan meningkatkan kinerja klasifikasi. Kombinasi CD-KNN dengan ANOVA dan SMOTE menghasilkan nilai akurasi, presisi, recall, F1-score, dan nilai ROC-AUC yang paling optimal. Hal ini membuktikan bahwa pemilihan fitur yang relevan dan penanganan data tidak seimbang berperan penting dalam meningkatkan efektivitas sistem deteksi. Oleh karena itu, pendekatan ini dapat menjadi alternatif yang efisien dan andal dalam sistem keamanan jaringan untuk mendeteksi serangan DDoS secara dini.

DAFTAR PUSTAKA

Alassaf, M., & Qamar, A. M. (2022). Improving Sentiment Analysis Of Arabic Tweets By One-Way ANOVA. *Journal of King Saud University - Computer and Information Sciences*, *34*(6), 2849–2859. https://doi.org/10.1016/j.jksuci.2020.10.023

Chandrashekar, G., & Sahin, F. (2014). A Survey On Feature Selection Methods. *Computers and Electrical Engineering*, *40*(1), 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. In *Journal of Artificial Intelligence Research* (Vol. 16).

Elreedy, D., Atiya, A. F., & Kamalov, F. (2024). A Theoretical Distribution Analysis Of Synthetic Minority Oversampling Technique (SMOTE) For Imbalanced Learning.

Machine Learning, *113*(7), 4903–4923. https://doi.org/10.1007/s10994-022-06296-4

Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. *IEEE Access*, *10*, 19083–19095. https://doi.org/10.1109/ACCESS.2022.3151048

Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based On SMOTE And Convolutional Neural Networks. *Applied Sciences (Switzerland)*, *13*(6). https://doi.org/10.3390/app13064006

Kumar, M., Rath, N. K., Swain, A., & Rath, S. K. (2015). Feature Selection And Classification Of Microarray Data Using Mapreduce Based ANOVA And K-Nearest Neighbor. *Procedia Computer Science*, *54*, 301–310. https://doi.org/10.1016/j.procs.2015.06.035

Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2021). Multiclass Confusion Matrix Reduction Method And Its Application On Net Promoter Score Classification Problem. *Technologies*, *9*(4). https://doi.org/10.3390/technologies9040081

Megantara, A. A., & Ahmad, T. (2020). Feature Importance Ranking For Increasing Performance Of Intrusion Detection System. *2020 3rd International Conference on Computer and Informatics Engineering, IC2IE 2020*, 37–42. https://doi.org/10.1109/IC2IE50715.2020.9274570

Muraina, I. O. (2022). *Ideal Dataset Splitting Ratios In Machine Learning Algorithms: General Concerns For Data Scientists And Data Analysts*. https://www.researchgate.net/publication/358284895

Musthafa, M. B., Huda, S., Kodera, Y., Ali, M. A., Araki, S., Mwaura, J., & Nogami, Y. (2024). Optimizing IoT Intrusion Detection Using Balanced Class Distribution, Feature Selection, And Ensemble Machine Learning Techniques. *Sensors*, *24*(13). https://doi.org/10.3390/s24134293

Qrator Labs. (2024). *Q3 2024 DDoS, Bots and BGP Incidents Statistics and Overview*.

Rácz, A., Bajusz, D., & Héberger, K. (2021). Effect Of Dataset Size And Train/Test Split Ratios In Qsar/Qspr Multiclass Classification. *Molecules*, *26*(4). https://doi.org/10.3390/molecules26041111

Ramli, N. E., Yahya, Z. R., & Said, N. A. (2022). Confusion Matrix As Performance Measure For Corner Detectors. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *29*(1), 256–265. https://doi.org/10.37934/araset.29.1.256265

Robindro, K., Singh, Y. R., Clinton, U. B., Takhellambam, L., & Hoque, N. (2022). CD-KNN: A Modified K-Nearest Neighbor Classifier With Dynamic K Value. *Lecture Notes in Electrical Engineering*, *925*, 753–762. https://doi.org/10.1007/978-981-19-4831-2_62

Salmi, S., & Oughdir, L. (2023). Performance Evaluation Of Deep Learning Techniques For Dos Attacks Detection In Wireless Sensor Network. *Journal of Big Data*, *10*(1). https://doi.org/10.1186/s40537-023-00692-w

Sharafaldin, I., Lashkari, A. H., Hakak, S., & Ghorbani, A. A. (2019). *Developing Realistic Distributed Denial Of Service (Ddos) Attack Dataset And Taxonomy*. IEEE.

SOC Radar. (2024). *Indonesia Threat Landscape Report*.

Varoquaux, G., & Colliot, O. (2023). Evaluating Machine Learning Models And Their Diagnostic Value. In *Neuromethods* (Vol. 197, pp. 601–630). Humana Press Inc. https://doi.org/10.1007/978-1-0716-3195-9_20

Wang, S., Dai, Y., Shen, J., & Xuan, J. (2021). Research On Expansion And Classification Of Imbalanced Data Based On SMOTE Algorithm. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-03430-5