Analisis Komparasi Prediksi Serangan DDoS Menggunakan Machine Learning

Sunardi[1]✉, Suyahman[2]
[1,2]Universitas Ahmad Dahlan, Indonesia

✉Corresponding Author: sunardi@mti.uad.ac.id

ABSTRAK

Serangan Distributed Denial of Service (DDoS) merupakan ancaman serius terhadap kestabilan jaringan komputer, yang dapat menyebabkan gangguan layanan pada sistem yang ditargetkan. Deteksi dini serangan DDoS sangat penting untuk meminimalkan dampaknya. Penelitian ini bertujuan untuk membandingkan kinerja empat algoritma machine learning—K-Nearest Neighbors (KNN), Random Forest (RF), XGBoost, dan Logistic Regression (LR)—dalam mendeteksi serangan DDoS menggunakan dataset besar dari Kaggle, yang terdiri dari lebih dari satu juta baris data. Dataset mencakup berbagai informasi lalu lintas jaringan, termasuk alamat IP sumber dan tujuan, port, panjang paket, serta jumlah paket yang dikirim dalam interval waktu tertentu. Hasil penelitian menunjukkan bahwa algoritma RF memiliki akurasi tertinggi, yaitu 99.83%, diikuti oleh KNN dan XGBoost dengan akurasi masing-masing 99.82%. Logistic Regression, meskipun menunjukkan akurasi yang baik sebesar 98.80%, memiliki performa yang lebih rendah dibandingkan algoritma lainnya karena keterbatasannya dalam menangani pola data non-linier. Berdasarkan hasil ini, dapat disimpulkan bahwa RF, KNN, dan XGBoost sangat efektif untuk deteksi serangan DDoS, sementara LR dapat digunakan sebagai alternatif dengan efisiensi komputasi yang lebih tinggi. Penelitian ini memberikan kontribusi signifikan terhadap pengembangan sistem deteksi DDoS berbasis machine learning yang lebih akurat dan efisien.

Kata kunci : DDoS, machine learning, deep learning, XGBoost, KNN

A. Pendahuluan

Serangan *Distributed Denial of Service* (DDoS) merupakan ancaman serius terhadap keamanan jaringan di era digital saat ini [1]. DDoS adalah jenis serangan yang bertujuan untuk membuat sumber daya atau layanan jaringan tidak dapat diakses oleh pengguna yang sah dengan membanjiri server atau jaringan dengan volume lalu lintas data yang sangat tinggi [2]. Seiring dengan berkembangnya teknologi, serangan DDoS semakin canggih dan sulit dideteksi, memaksa ahli keamanan untuk terus mencari solusi yang lebih efektif dalam mendeteksi dan mencegah serangan tersebut [3].

Dalam beberapa tahun terakhir, pendekatan berbasis *machine learning* (ML) telah mendapatkan perhatian yang besar dalam deteksi dan prediksi serangan DDoS [4]. Algoritma machine learning memiliki kemampuan untuk mempelajari pola serangan dari data historis dan mengidentifikasi anomali dalam lalu lintas jaringan, yang membuatnya sangat berguna dalam mendeteksi serangan DDoS secara dini [5]. Berbagai algoritma *machine learning*, seperti K-Nearest Neighbors (KNN), Random Forest (RF), XGBoost, dan Logistic Regression (LR), telah diuji untuk efektivitasnya dalam memprediksi pola data [6]-[9].

Salah satu penelitian yang relevan dilakukan oleh Rahman et al. (2024), yang memanfaatkan model machine learning Multilayer Perceptron Classifier (MLPC) untuk membedakan antara lalu lintas yang aman (benign) dan berbahaya (malicious) [10]. Dataset yang digunakan adalah DDoS Attack SDN Dataset, yang terdiri dari 23 fitur dengan total 104,345 data. Hasil uji coba menunjukkan bahwa model MLPC mencapai tingkat akurasi sebesar 99,05%, dengan precision dan recall masing-masing mencapai 99% dalam mendeteksi lalu lintas yang aman maupun berbahaya. Penelitian ini membuktikan bahwa model MLPC

dapat diandalkan dalam klasifikasi lalu lintas jaringan dan mendeteksi serangan DDoS dengan tingkat akurasi yang tinggi.

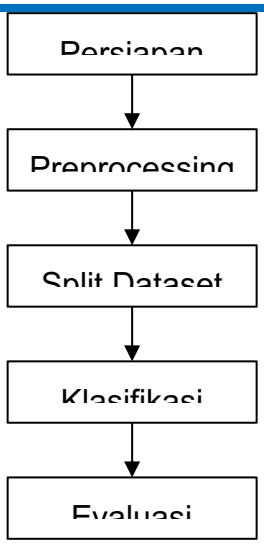
Beberapa pendekatan berbasis machine learning lainnya juga menawarkan solusi yang menjanjikan dalam mendeteksi serangan DDoS dengan lebih efisien. Dengan memanfaatkan teknik analisis data yang canggih, model machine learning dapat belajar dari pola lalu lintas jaringan dan mengidentifikasi aktivitas mencurigakan secara real-time. Misalnya, Perez-Diaz et al. (2021) menggunakan algoritma Random Tree, SVM, Random Forest, REP Tree, J48, dan MLP dalam mendeteksi LR-DDoS Attacks, dengan akurasi yang dicapai mencapai 95% [11]. Ye et al. (2020) menggunakan metode Support Vector Machine (SVM) untuk mendeteksi serangan DDoS pada SDN, dengan akurasi 95,24% [12]. Selain itu, Tan et al. (2021) mengembangkan metode hybrid dengan menggabungkan K-Mean dan K-Nearest Neighbors (KNN), mencapai tingkat akurasi sebesar 98,85% [13]. Meskipun hasil-hasil tersebut cukup baik, namun model-model tersebut masih menghadapi tantangan dalam mencapai performa yang maksimal, terutama dalam mendeteksi serangan DDoS yang semakin kompleks.

Untuk mendalami lebih lanjut deteksi serangan DDoS, penelitian ini menggunakan dataset dari Kaggle yang mengandung berbagai data jaringan untuk menguji dampak serangan DDoS terhadap lalu lintas jaringan. Dataset ini terdiri dari 1.048.444 baris data, yang mencakup informasi komunikasi dari berbagai lapisan jaringan, alamat IP sumber dan tujuan, nomor port, panjang paket, jumlah paket yang dikirim, dan interval waktu. Dataset ini menyediakan informasi yang sangat rinci mengenai lalu lintas jaringan yang dapat digunakan untuk membedakan antara lalu lintas yang normal dan lalu lintas yang disebabkan oleh serangan DDoS.

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja keempat algoritma tersebut—K-Nearest Neighbors (KNN), Random Forest (RF), XGBoost, dan Logistic Regression (LR)—dalam memprediksi serangan DDoS berdasarkan akurasi dan kemampuannya dalam mengidentifikasi serangan. Dengan menggunakan dataset yang relevan, penelitian ini diharapkan dapat memberikan wawasan yang lebih dalam mengenai efektivitas berbagai algoritma machine learning dalam mengatasi ancaman DDoS, serta memberikan rekomendasi yang dapat diimplementasikan dalam sistem deteksi serangan yang lebih baik di dunia nyata.

B. Metode

Pada penelitian ini, tahapan yang dilakukan untuk memprediksi serangan Distributed Denial of Service (DDoS) melibatkan beberapa langkah utama, mulai dari preprocessing data, pembagian dataset, klasifikasi menggunakan berbagai algoritma machine learning, hingga evaluasi menggunakan confusion matrix. Gambar 1. menunjukan tahapan-tahapan dalam penelitian ini secara keseluruhan, mulai dari tahap preprocessing hingga evaluasi model.



Gambar 1. Tahapan Penelitian

1. Dataset

Dataset yang digunakan dalam penelitian ini berasal dari Kaggle dan mencakup data jaringan yang bertujuan untuk menguji dampak serangan DDoS terhadap lalu lintas jaringan (https://www.kaggle.com/datasets/oktayrdeki/ddos-traffic-dataset/). Dataset ini terdiri dari 1.048.444 baris data, yang mencakup berbagai informasi terkait komunikasi di berbagai lapisan jaringan. Beberapa kolom utama dalam dataset meliputi Highest Layer, Transport Layer, Source IP, Dest IP, Source Port, Dest Port, Packet Length, Packets/Time, dan Target. Kolom Target menunjukkan apakah setiap baris data berhubungan dengan serangan DDoS atau lalu lintas normal, dengan nilai 0 untuk lalu lintas normal dan 1 untuk serangan DDoS. Contoh dataset bisa dilihat pada Gambar 2.

	Highest Layer	Transport Layer	Dest IP	Source Port	Dest Port	Packet Length	Packets/Time	target
0	ARP	UDP	192.168.1.1	0	0	60	92.8	1
1	ARP	UDP	192.168.1.10	0	0	42	92.9	0
2	ARP	UDP	192.168.1.1	0	0	60	362.8	1
3	ARP	UDP	192.168.1.12	0	0	42	362.8	0
4	ARP	UDP	192.168.1.1	0	0	60	364.3	1

Gambar 2. Contoh Dataset

2. Preprocessing

Tahapan pertama dalam penelitian ini adalah preprocessing data [14]. Sebelum data dapat digunakan untuk pelatihan model machine learning, kolom-kolom kategorikal dalam dataset diubah menjadi format numerik menggunakan metode Label Encoding [15]. Kolom seperti Highest Layer, Transport Layer, Source IP, dan Dest IP diubah menjadi representasi numerik untuk memudahkan pemrosesan lebih lanjut. Proses ini juga mencakup normalisasi untuk fitur numerik seperti Packet Length dan Packets/Time, guna memastikan agar data berada dalam rentang yang serupa dan memfasilitasi algoritma machine learning dalam menemukan pola yang lebih akurat [16].

3. Split Dataset

Setelah preprocessing selesai, data dibagi menjadi dua bagian: 70% untuk pelatihan model dan 30% untuk pengujian [17]. Pembagian ini bertujuan untuk menguji kemampuan model

dalam mengenali pola dan generalisasi terhadap data yang belum pernah dilihat sebelumnya [18]. Pembagian ini memungkinkan model untuk dilatih dengan data yang cukup, sementara data uji digunakan untuk mengevaluasi kinerja model yang telah dilatih. Dengan membagi data menjadi dua set ini, model dapat diuji secara objektif untuk mengetahui sejauh mana kemampuannya dalam mendeteksi serangan DDoS pada data yang tidak dikenal [19].

4. Klasifikasi

Setelah dataset dibagi, penelitian ini melanjutkan ke tahap klasifikasi dengan menggunakan empat algoritma machine learning, yaitu K-Nearest Neighbors (KNN), Random Forest (RF), XGBoost, dan Logistic Regression (LR).

K-Nearest Neighbors (KNN) adalah algoritma klasifikasi berbasis instance yang mengklasifikasikan data baru berdasarkan mayoritas kelas dari k tetangga terdekat dalam ruang fitur [20]. Jarak antar titik data dihitung menggunakan metrik seperti Euclidean distance. Algoritma ini sangat sederhana dan efektif, namun bisa menjadi kurang efisien pada dataset besar karena membutuhkan waktu komputasi tinggi untuk mencari tetangga terdekat.

Random Forest, di sisi lain, adalah algoritma ensemble yang menggabungkan banyak pohon keputusan (decision trees) [21]. Setiap pohon dilatih pada subset acak dari data pelatihan dengan pemilihan fitur secara acak, dan hasil akhirnya diperoleh melalui voting mayoritas untuk klasifikasi. Keunggulannya adalah kemampuannya dalam menghindari overfitting dan memberikan hasil yang stabil meskipun data mengandung noise.

XGBoost (Extreme Gradient Boosting) adalah algoritma berbasis gradient boosting yang membangun model prediksi secara berurutan, dengan setiap model baru bertujuan untuk memperbaiki kesalahan dari model sebelumnya [22]. XGBoost terkenal karena kemampuannya menangani dataset besar dan kompleks dengan efisiensi tinggi, serta fleksibilitas dalam tuning parameter untuk meningkatkan performa. Terakhir, Logistic Regression adalah algoritma klasifikasi yang menggunakan fungsi logistik (sigmoid) untuk memodelkan probabilitas terjadinya suatu kejadian. Meskipun disebut regresi, algoritma ini digunakan untuk klasifikasi biner, seperti dalam mendeteksi serangan DDoS.

Logistic Regression efektif pada dataset dengan fitur yang linier, namun kurang optimal untuk masalah yang memiliki hubungan non-linier antar fitur [23]. Keempat algoritma ini dipilih dalam penelitian ini untuk membandingkan kinerjanya dalam mendeteksi serangan DDoS pada dataset yang digunakan.

5. Evaluasi

Tahap terakhir dalam penelitian ini adalah evaluasi model. Hasil prediksi dari masing-masing model dievaluasi menggunakan confusion matrix yang menggambarkan jumlah prediksi yang benar dan salah [24]. Metrik seperti akurasi, precision, recall, dan F1-score dihitung berdasarkan confusion matrix untuk menilai seberapa baik setiap model dalam mendeteksi serangan DDoS [25]. Dalam hal ini, evaluasi ini tidak hanya mengukur akurasi model, tetapi juga mengidentifikasi keseimbangan antara prediksi yang benar dan salah, terutama dalam konteks deteksi serangan yang mungkin terdistribusi secara tidak merata.

Dengan melakukan analisis terhadap hasil evaluasi ini, perbandingan antara algoritma-algoritma yang digunakan dapat dilakukan untuk menentukan model mana yang paling efektif dalam mendeteksi serangan DDoS pada dataset yang ada.

C. Hasil dan Pembahasan

Pada penelitian ini, empat algoritma machine learning, yaitu K-Nearest Neighbors (KNN), Random Forest (RF), XGBoost, dan Logistic Regression (LR), telah diterapkan untuk mendeteksi serangan DDoS pada dataset yang diambil dari Kaggle. Hasil uji coba yang

diperoleh menunjukkan bahwa masing-masing algoritma menghasilkan tingkat akurasi yang sangat baik, dengan sedikit perbedaan di antara mereka.

Tabel 1. Perbandingan Kinerja KNN, Random Forest, XGBoost, dan Logistic Regression

Algoritma	Akurasi	Presisi	Recall	F1-Score
KNN	0.9982	0.9982	0.9982	0.9982
Random Forest	0.9983	0.9983	0.9983	0.9983
XGBoost	0.9982	0.9982	0.9982	0.9982
Logistic Regression	0.9880	0.9883	0.9876	0.9880

Hasil penelitian pada Tabel 1. menunjukkan bahwa algoritma Random Forest (RF) memiliki performa terbaik dalam mendeteksi serangan DDoS dibandingkan dengan algoritma lainnya. Berdasarkan hasil evaluasi, RF mencapai nilai akurasi, presisi, recall, dan F1-score sebesar 99.83%, yang menandakan kemampuannya dalam mengidentifikasi pola lalu lintas jaringan yang mencurigakan dengan sangat akurat. Algoritma K-Nearest Neighbors (KNN) dan XGBoost juga menunjukkan performa yang hampir setara dengan RF, masing-masing memperoleh nilai akurasi, presisi, recall, dan F1-score sebesar 99.82%. Hal ini mengindikasikan bahwa ketiga algoritma tersebut mampu menangkap karakteristik data yang relevan dalam mendeteksi serangan DDoS secara konsisten.

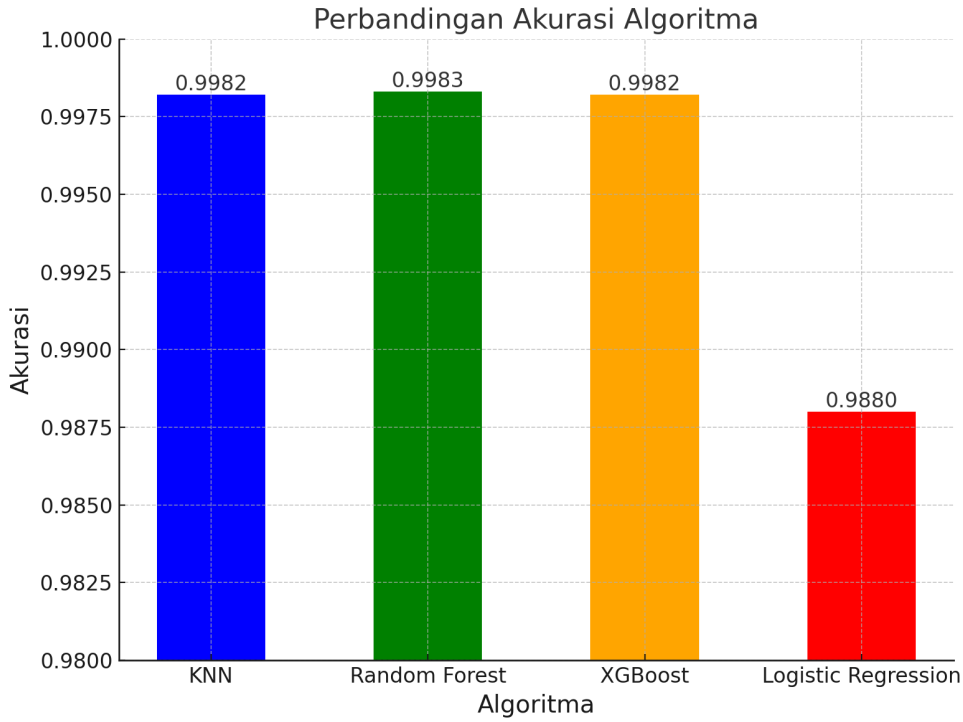
Sementara itu, Logistic Regression (LR) menunjukkan hasil yang lebih rendah dibandingkan algoritma lainnya, dengan akurasi 98.80%, presisi 98.83%, recall 98.76%, dan F1-score 98.80%. Performa LR yang sedikit tertinggal dapat disebabkan oleh keterbatasannya dalam menangani data yang memiliki pola non-linear, yang sering muncul dalam lalu lintas jaringan yang kompleks. Namun, nilai akurasi mendekati 99% menunjukkan bahwa LR tetap memiliki potensi sebagai metode deteksi, terutama dalam skenario di mana efisiensi komputasi lebih diutamakan.

Metrik evaluasi yang digunakan, seperti presisi, recall, dan F1-score, memberikan wawasan penting untuk memahami performa setiap algoritma. Presisi yang tinggi pada ketiga algoritma utama (RF, KNN, dan XGBoost) menunjukkan kemampuan mereka dalam mengurangi kesalahan positif dengan memastikan bahwa prediksi serangan DDoS benar-benar akurat. Recall yang tinggi, di sisi lain, mencerminkan kemampuan algoritma untuk mendeteksi sebanyak mungkin serangan yang ada dalam data, sehingga mengurangi risiko serangan tidak terdeteksi. Harmonisasi antara presisi dan recall yang dicerminkan oleh F1-score menguatkan bukti bahwa ketiga algoritma tersebut sangat efektif untuk deteksi serangan DDoS.

Dataset yang digunakan dalam penelitian ini, dengan lebih dari satu juta baris data, mencakup berbagai informasi lalu lintas jaringan seperti panjang paket, jumlah paket per waktu, serta alamat IP sumber dan tujuan. Kompleksitas dataset ini memungkinkan algoritma untuk belajar dari pola yang merepresentasikan aktivitas serangan. Algoritma RF, KNN, dan XGBoost mampu memanfaatkan informasi ini dengan optimal, sementara LR menunjukkan keterbatasan dalam menangkap pola yang lebih kompleks.

Dari analisis ini, dapat disimpulkan bahwa Random Forest adalah algoritma yang paling unggul untuk mendeteksi serangan DDoS, dengan performa yang sedikit lebih baik dibandingkan KNN dan XGBoost. Namun, kedua algoritma terakhir juga sangat kompetitif dan dapat digunakan sebagai alternatif dalam implementasi sistem deteksi DDoS. Logistic Regression, meskipun kurang optimal dalam menangani data yang kompleks, tetap

memberikan hasil yang cukup baik dan dapat menjadi pilihan dalam situasi dengan keterbatasan sumber daya komputasi. Temuan ini memberikan dasar yang kuat untuk pengembangan sistem deteksi serangan DDoS berbasis machine learning yang lebih efektif dan efisien. Gambar 3. menunjukkan perbandingan akurasi antara keempat algoritma yang diuji dalam penelitian ini.



Gambar 3. Perbandingan Akurasi KNN, Random Forest, XGBoost, dan Logistic Regression

Dari hasil tersebut, dapat disimpulkan bahwa Random Forest, KNN, dan XGBoost memiliki performa yang sangat baik dalam mendeteksi serangan DDoS dengan akurasi yang hampir identik, yaitu di atas 99.82%. Random Forest sedikit lebih unggul dibandingkan KNN dan XGBoost dengan akurasi tertinggi, yaitu 99.83%. Meskipun Logistic Regression juga menunjukkan hasil yang cukup baik dengan akurasi mendekati 99%, algoritma ini sedikit tertinggal karena keterbatasannya dalam menangani data dengan hubungan non-linier yang sering terjadi pada pola lalu lintas jaringan. Pemilihan algoritma terbaik akan sangat bergantung pada karakteristik data dan kebutuhan spesifik dalam mendeteksi serangan DDoS, seperti efisiensi komputasi atau kemampuan menangani data yang lebih kompleks.

D. Simpulan

Penelitian ini bertujuan untuk membandingkan kinerja empat algoritma machine learning—K-Nearest Neighbors (KNN), Random Forest (RF), XGBoost, dan Logistic Regression (LR)—dalam mendeteksi serangan DDoS menggunakan dataset yang berisi lebih dari satu juta baris data yang merepresentasikan pola lalu lintas jaringan. Hasil uji coba menunjukkan bahwa algoritma Random Forest mencapai performa terbaik dengan akurasi 99.83%, diikuti oleh KNN dan XGBoost yang masing-masing memiliki akurasi 99.82%. Ketiga algoritma ini menunjukkan efektivitas yang tinggi dalam mendeteksi serangan DDoS pada dataset besar. Logistic Regression, meskipun memiliki akurasi yang sedikit lebih rendah, yaitu 98.80%, tetap menunjukkan performa yang baik. Penurunan kinerja ini dapat disebabkan oleh keterbatasannya dalam menangani pola data yang bersifat non-linier. Secara keseluruhan, Random Forest, KNN, dan XGBoost menunjukkan performa yang hampir sebanding dan

sangat efektif untuk mendeteksi serangan DDoS, menjadikannya pilihan utama untuk implementasi deteksi berbasis machine learning. Namun, pemilihan algoritma terbaik juga harus mempertimbangkan faktor lain, seperti efisiensi komputasi, yang dapat menjadi penting dalam lingkungan dengan keterbatasan sumber daya. Penelitian ini memberikan dasar yang kuat untuk pengembangan model deteksi serangan DDoS yang lebih efisien dan akurat, dengan menyesuaikan algoritma yang digunakan berdasarkan karakteristik dataset dan kebutuhan spesifik sistem deteksi.

Daftar Pustaka

[1] Yan, Q., Yu, F. R., Gong, Q., & Li, J. (2015). Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: A survey, some research issues, and challenges. *IEEE Communications Surveys & Tutorials*, 18(1), 602-622.

[2] Douligeris, C., & Mitrokotsa, A. (2004). DDoS attacks and defense mechanisms: classification and state-of-the-art. Computer Networks, 44(5), 643-666.

[3] Kumari, P., & Jain, A. K. (2023). A comprehensive study of DDoS attacks over IoT network and their countermeasures. *Computers & Security*, 127, 103096.

[4] Ali, T. E., Chong, Y. W., & Manickam, S. (2023). Machine learning techniques to detect a DDoS attack in SDN: A systematic review. Applied Sciences, 13(5), 3183.

[5] Al-Shareeda, M. A., Manickam, S., & Saare, M. A. (2023). DDoS attacks detection using machine learning and deep learning techniques: Analysis and comparison. *Bulletin of Electrical Engineering and Informatics*, 12(2), 930-939.

[6] Kramer, O., & Kramer, O. (2013). K-nearest neighbors. In *Dimensionality Reduction with Unsupervised Nearest Neighbors* (pp. 13-23).

[7] Parmar, A., Katariya, R., & Patel, V. (2019). A review on random forest: An ensemble classifier. In *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018* (pp. 758-763). Springer International Publishing.

[8] Aydin, Z. E., & Ozturk, Z. K. (2021, March). Performance analysis of XGBoost classifier with missing data. In *1st International Conference on Computing and Machine Intelligence*, no. 1, pp. [page numbers]. (Note: The page numbers are missing in the provided citation. Please replace [page numbers] with the actual page range if available.)

[9] Nick, T. G., & Campbell, K. M. (2007). Logistic regression. In *Topics in Biostatistics* (pp. 273-301).

[10] H. RAHMAN and Tata Sutabri, "Analisis Serangan DDoS Menggunakan Machine Learning Pada Arsitektur Software-Defined Network," JSAI (Journal Scientific and Applied Informatics), vol. 7, no. 3, pp. 531–536, Nov. 2024, doi: https://doi.org/10.36085/jsai.v7i3.7301.

[11] J. A. Perez-Diaz, I. A. Valdovinos, K. K. R. Choo, and D. Zhu, "A Flexible SDN-Based Architecture for Identifying and Mitigating Low-Rate DDoS Attacks Using Machine Learning," IEEE Access, vol. 8, pp. 155859–155872, 2020, doi: 10.1109/ACCESS.2020.3019330.

[12] J. Ye, X. Cheng, J. Zhu, L. Feng, and L. Song, "A DDoS Attack Detection Method Based on SVM in Software Defined Network," Security and Communication Networks, vol. 2018, no. 1, p. 9804061, Jan. 2018, doi: 10.1155/2018/9804061.

[13] L. Tan, Y. Pan, J. Wu, J. Zhou, H. Jiang, and Y. Deng, "A New Framework for DDoS Attack Detection and Defense in SDN Environment," IEEE Access, vol. 8, pp. 161908–161919, 2020, doi: 10.1109/ACCESS.2020.3021435.

[14] García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. *Big Data Analytics*, 1, 1-22.
[15] Herdian, C., Kamila, A., & Budidarma, I. G. A. M. (2024). Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi. Technologia: Jurnal Ilmiah, 15(1), 93-108.
[16] Deekshith, A. (2021). Data engineering for AI: Optimizing data quality and accessibility for machine learning models. *International Journal of Management Education for Sustainable Development*, 4(4), 1-33.
[17] Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., ... & Pham, B. T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 2021(1), 4832864.
[18] Kuncheva, L. I. (2014). Combining pattern classifiers: methods and algorithms. John Wiley & Sons.
[19] Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7, 1-8.
[20] Steinbach, M., & Tan, P. N. (2009). kNN: k-nearest neighbors. In The top ten algorithms in data mining (pp. 165-176). Chapman and Hall/CRC.
[21] Shaik, A. B., & Srinivasan, S. (2019). A brief survey on random forest ensembles in classification model. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, Volume 2 (pp. 253-260). Springer Singapore.
[22] Aksoy, N., & Genc, I. (2023). Predictive models development using gradient boosting based methods for solar power plants. Journal of Computational Science, 67, 101958.
[23] Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11), 1225-1231.
[24] S. Suyahman, S. Sunardi, M. Murinto, and A. N. Khusna, “Data Augmentation Using Test-Time Augmentation on Convolutional Neural Network-Based Brand Logo Trademark Detection,” Indonesian Journal of Artificial Intelligence and Data Mining, vol. 7, no. 2, p. 266, Apr. 2024, doi: https://doi.org/10.24014/ijaidm.v7i2.28804.
[25] Suyahman, Sunardi, and Murinto, “Comparative Analysis of CNN Architectures in Siamese Networks with Test-Time Augmentation for Trademark Image Similarity Detection,” Scientific Journal of Informatics, vol. 11, no. 4, pp. 949–958, Dec. 2024, doi: https://doi.org/10.15294/sji.v11i4.13811.