Identifikasi Serangan *Lowrate Distributed Denial Of Services* Dalam Jaringan Dengan Menggunakan Algoritma Adaboost

Muhammad Ilman Aqilaa [1], Diash Firdaus[2], Nawaf Naofal[3]

[1,2,3]Universitas Logistik Bisnis Internasional
Jalan Sariasih Nomor. 54 Sarijadi, Kota Bandung, Jawa Barat, Indonesia
e-mail: [1]ilmanaqilaa2@gmail.com, [2]diashfirdaus@gmail.com, [3]nawafnaofal@gmail.com,

Artikel Info : Diterima : 16-02-2023 | Direvisi : 19-06-2023 | Disetujui : 28-06-2023

Abstrak - Distributed Deniel of Service (DDoS) merupakan salah satu serangan yang paling populer saat ini. DDoS adalah yang bertujuan untuk menyebabkan crash pada sistem server dengan cara membanjiri paket ataupun permintaan pada jaringan. Karakteristik serangan *lowrate Distributed Denial of Service* (DDoS) sulit di bedakan dari arus lalu lintas jaringan normal, sehingga untuk mengidentifikasi serangan ini diperlukan sistem yang dapat mengklasifikasi serangan *lowrate* DDoS. Pada penelitian ini telah dilakukan identifikasi serangan *lowrate* DDoS dengan menggunakan metode *AdaBoost*. Dataset yang digunakan merupakan data *MachineLearningCSV* yaitu bagian dari kumpulan data CICIDS-2017 yang berasal dari *Konsorsium* ISCX. *MachineLeaningCSV* terdiri dari delapan (8) sesi pemantauan *traffic*. Data tersebut akan di preprocessing guna mengubah data menjadi *array* dan dilakukan seleksi atribut yang paling relevan untuk mempermudah kinerja metode AdaBoost dalam melakukan klasifikasi. Setelah melakukan pengujian terhadap penlitian terhadap deteksi DdoS dengan menggunakan algoritma AdaBoost, akurasi yang dihasilkan dengan tingkat akurasi hingga 99.2%.

Kata Kunci : DdoS, AdaBoost, *Machine Learning, CICIDS-2017*

Abstracts *- Distributed Deniel of Service (DDoS) is one of the most popular attacks today. DDoS attacks aim to crash a server system by flooding packets or requests on the network. The characteristics of lowrate Distributed Denial of Service (DDoS) attacks are difficult to distinguish from normal network traffic flow, so to identify this attack a system that can classify lowrate DDoS attacks is needed. In this research, identification of DDoS attacks has been carried out using the AdaBoost method. The dataset used is MachineLearningCSV data, which is part of the CICIDS-2017 dataset originating from the ISCX Consortium. MachineLeaningCSV consists of eight (8) traffic monitoring traffic. Data will be preprocessed to convert the data into an array and select the most relevant attributes to facilitate the performance of the AdaBoost method in classification. After testing research on DdoS detection using the AdaBoost algorithm, the resulting accuracy is up to 99.2%.*

Keywords : DdoS, AdaBoost, *Machine Learning, CICIDS-2017*

1. PENDAHULUAN

Penggunaan teknologi saat ini sangat berkembang pesat dengan nilai guna yang sangat tinggi sehingga dapat membantu meringankan setiap kegiatan yang dilakukan, ditambah dengan jaringan yang dapat menghubungkan perangkat satu dengan yang lainnya. Namun dalam satu waktu kegiatan tersebut dapat menghambat dengan adanya serangan melalui jaringan yaitu *Distributed Denial of Service* atau biasa di sebut DdoS. Serangan DDoS memiliki efek yang serius terhadap suatu perusahaan dan mengakibatkan kerugian yang sangat besar. Cara kerja serangan ini yaitu dengan mengirimkan request terhadap sebuah komputer atau server dalam jumlah yang melebihi kemampuan komputer itu. Serangan DDoS sendiri memiliki beberapa jenis, diantaranya *SYN Flood*, *TCP Anomaly*, *DNS Flood*, *UDP Flood*, *ACK Flood*, dan lain sebagainya (Deepa et al., 2018)

Distributed Denial of Service (DDOS) merupakan aktifitas pengiriman paket dalam jaringan dalam jumlah besar yang ditujukan untuk membanjiri jaringan dengan data sehingga suatu host menjadi tidak dapat diakses oleh pengguna yang berhak (Yang & Zhao, 2019) DDoS (Distributed Denial of Service) adalah jenis serangan yang menggunakan banyak "Zombie" PC yang masing-masing melakukan serangan DoS (Denial of Service) ke target yang sama. Serangan ini bertujuan untuk melumpuhkan target dengan cara membanjiri jalur data dengan paket-

paket illegal. DdoS lebih terstruktur dan juga dampak yang dihasilkan lebih besar karena jumlah penyeranglebih banyak. DDoS attack akan menguras sumber daya korban seperti CPU, disk space, maupun bandwidth sehingga komputer atau system tidak dapat berfungsi dengan maksimal (Sumadi & Aditya, 2021), (Ashfaq et al., 2022)

Machine learning merupakan serangkaian teknik yang dapat membantu dalam menangani dan memprediksi data yang sangat besar dengan cara mempresentasikan data-data tersebut dengan algoritma pembelajaran (*Implementasi Deep Learning Menggunakan Convolutional Neural Network Untuk Klasifikasi Citra Candi Berbasis GPU*, 2017). Menurut (Khuphiran et al., 2018) machine learning dapat didefinisikan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat. Definisi pengalaman disini ialah informasi sebelumnya yang telah tersedia dan bisa dijadikan data pembelajar.

Maka dari itu akan dibuatkan deteksi dengan menghitung akurasi kemungkinan yang terjadi untuk menghindari serangan DDoS dengan menggunakan Machine Learning yang mengolah dataset sehingga akan menghasilkan akurasi dari dataset tersebut menggunakan algoritma AdaBoost (Yadahalli & Nighot, 2018). Algoritma AdaBoost digunakan pada penelitian ini dikarenakan akan menghasilkan akurasi yang tinggi. Dan untuk dataset yang akan digunakan dalam deteksi yaitu dataset CICIDS-2017 (Shailesh Singh Panwar et al., 2022), alasan pemilihan dataset tersebut dikarenakan mewakili *traffiuc* jaringan dunia nyata dalam percobaan.

2. LANDASAN TEORI

2.1. AdaBoost

AdaBoost adalah metode pembelajaran ensemble yang terkenal yang menyediakan strategi yang efektif untuk menghasilkan pembelajar yang kuat dengan melatih individu secara iteratif individu pembelajar. AdaBoost awalnya digunakan untuk mengintegrasikan algoritma klasifikasi dan algoritma pohon regresi. AdaBoost mengimplementasikan prediksi ensemble sebagai prediksi ensembel. AdaBoost menerapkan prediksi ensemble sebagai langkah-langkah berikut: (1) dataset pelatihan didistribusikan menggunakan bobot awal; (2) base-learner dilatih berdasarkan dataset tertimbang; (3) bobot dari base-learner dan dataset akan diperbarui base-learner dan dataset akan diperbarui sesuai dengan kinerja base-learner pada putaran iterasi sebelumnya; (4) iterasi sebelumnya; (4) menghentikan iterasi jika kondisi kondisi terminasi terpenuhi; (5) menggabungkan baselearner menggunakan weighted voting. Bobot iteratif memungkinkan pembelajar berikutnya untuk fokus lebih pada kasus-kasus yang salah klasifikasi. Dengan demikian, kombinasi dari pembelajar dasar dapat mengoreksi bias klasifikasi dan cenderung menghasilkan performa yang baik (Hu, 2017). AdaBoost (Adaptive Boosting) adalah Family Boosting dari algoritma algoritma Boosting. Jenis ini perhatian lebih pada sampel yang dikategorikan salah selama pelatihan, menyesuaikan distribusi sampel, dan mengulangi operasi ini hingga jumlah pelatihan pengklasifikasi nilai yang telah ditentukan sebelumnya, menyelesaikan pembelajaran dan keluar dari loop

2.2.Dataset CICIDS-2017

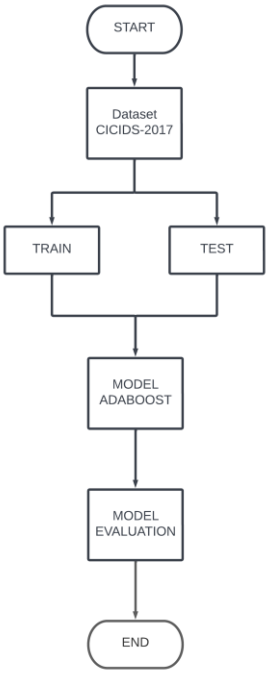
Dataset yang digunakan merupakan data *MachineLearningCSV* yaitu bagian dari kumpulan data CICIDS-2017 yang berasal dari *Konsorsium* ISCX. *MachineLeaningCSV* terdiri dari delapan (8) sesi pemantauan *traffic*, dimana masing-masing data tersebut berbentuk koma yang dipisahkan oleh suatu nilai (CSV). File tersebut berisikan *traffic* normal sebagai "*Benign Traffic*" dan "*Traffic Attacks*". *Traffic* serangan memiliki lebih banyak detail di dalam kolom pada tabel ke dua (2). Selain *Benign Traffic* dan *Attacks Traffic* terdapat empat belas (14) serangan dalam dataset ini. Pada karya yang dibuat, penulis mempertimbangkan fitur kompleks itu merupakan serangan canggih berbasis jaringan secara modern pada atribut *traffic* (Kurniabudi et al., 2020)

2.3. Virtual Machine

Virtual machine adalah program yang berguna untuk melakukan simulasi suatu sistem PC lengkap. Yang dimaksud lengkap di sini adalah RAM, hard disk, floppy disk, prosesor, graphics card dan beberapa device lain yang umumnya terdapat pada PC. Program semacam ini mungkin tidak banyak berguna bagi sebagian orang, tapi untuk kebutuhan tertentu atau spesifik, manfaatnya akan sangat terasa, simulasi penerapan Mikrotik Router sebagai user manager dalam sebuah jaringan internet hostpot (Amarudin & Yuliansyah, 2018). VirtualBox merupakan program open source terkait virtualisasi. Virtualisasi yaitu teknologi yang memungkinkan untuk membangun komputer PC virtual yang dapat berfungsi secara independen dari sistem operasi. Komputer host mensimulasikan semua jenis perangkat keras yang terkait dengan mesin virtual. Jika seseorang ingin menguji dan meniru instalasi sistem tanpa kehilangan sistem yang ada, kemampuan ini sangat penting. Tampaknya kita dapat memiliki beberapa jenis perangkat PC dengan beberapa sistem operasi yang memanfaatkan VirtualBox tanpa harus memiliki peralatan yang sebenarnya (Alfidzar & Zen, 2022).

3. METODE PENELITIAN

3.1. Kerangka Penelitian



Gambar 1 Kerangka Penelitian

Berdasarkan gambar diatas, akan dijelaskan tahapan-tahapan yang dilakukan dalam penilitian ini adalah sebagai berikut:

- Memulai dengan dataset CICIDS-2017 sebagai penggunaan data yang akan digunakan ke dalam machine learning untuk diolah hingga mendapatkan akurasi.
- Melakukan proses *Train* dan *Test* terhadap dataset CICIDS-2017 dengan menggunakan library *sklearn*. Mendefinisikan data menjadi source dan target. Source yang digunakan menyesuaikan dengan kolom yang tersedia pada dataset sebagai "X" terkecuali kolom terakhir yang berada pada ujung sebelah kanan yaitu kolom "*Label*" sebagai "Y". X dan Y adalah nama variabel yang digunakan saat mendefinisikan data source dan data target. Menjalankan proses train dan set yang nantinya akan dilanjutkan ke fase 3 yaitu penentuan model dan hasil akurasi. Dataset CICIDS2017 dibagi menjadi dua dengan 80% sebagai data pelatihan (train) dan 20% untuk pengujian (test).
- Proses model akan menggunakan algoritma *AdaBoost* untuk mendapatkan hasil akurasi yang berasal dari dataset yang sudah diproses pada proses sebelumnya. Menjalankan model tersebut sehingga akurasi didapatkan. Akurasi dataset CICIDS-2017 mendapatkan nilai sebesar 99.17%.

3.2. Evaluasi Model

Proses selanjutnya dilanjutkan dengan melakukan perhitungan untuk menampilkan klasifikasi data dengan menggunakan Confusion Matrix. Confusion matrix adalah salah satu cara yang sering digunakan untuk mengevaluasi kinerja model machine learning, termasuk model yang dibuat menggunakan algoritma decision tree. Confusion matrix adalah tabel yang menggambarkan hasil prediksi model dibandingkan dengan hasil yang sebenarnya.

Penggunaan confusion matrix diperlukan dua kelas yang akan diprediksi oleh model, biasanya disebut sebagai "positif" dan "negatif". Kemudian, kita akan menghitung jumlah true positive (TP), false positive (FP), true negative (TN), dan false negative (FN). Jumlah-jumlah tersebut akan digunakan untuk menghitung beberapa metrik yang umum digunakan untuk mengevaluasi kinerja model, seperti akurasi, presisi, dan recall (Firdaus et al., 2020).

Table 1 Tabel Configuration Matrix

	Hasil Serangan	Hasil Normal
Prediksi Serangan	TP (True Positive)	FP (False Positive)
Prediksi Normal	FN (False Negative)	TN (True Negative)

- *Accuracy* adalah persentase yang akan memberikan gambaran seberapa akurat model yang digunakan dalam mengklasifikasi dengan benar:

$$Accurracy = \frac{(TP+TN)}{(TP+FP+FN+TN)}$$

- Presisi adalah persentase yang menggambarkan akurasi antar data yang diminta dengan hasil prediksi yang diberikan oleh model:

$$Precision = \frac{(TP)}{(TP+FP)}$$

- *Recall* atau *sensitifity* adalah presentase yang menggambarkan berhasilnya suatu model dalam menemukan kembali sebuah informasi:

$$Recall = \frac{(TP)}{(TP+FN)}$$

- *F1 score* merupakan presentase perbandingan ratarata antara *precision* dan *recall* yang dibobotkan. *Accuracy* yang tepat akan digunakan sebagai acuan performansi algoritma jika dataset memiliki jumlah data False Negatif dan False Positif yang sangat mendekati.:

$$F1\ Score = \frac{(2*Recall*Precision)}{(Recall+Precision)}$$

Jumlah TP menunjukkan jumlah paket yang benar-benar merupakan serangan DDoS dan diprediksi dengan benar oleh model sebagai serangan. Jumlah FP menunjukkan jumlah paket yang normal namun salah diprediksi sebagai serangan oleh model. Jumlah FN menunjukkan jumlah paket yang merupakan serangan namun salah diprediksi sebagai normal oleh model. Jumlah TN menunjukkan jumlah paket yang benar-benar normal dan diprediksi dengan benar oleh model sebagai normal.

4. HASIL DAN PEMBAHASAN

Bab yang akan menampilkan kebutuhan dan hasil penelitian dengan hasil skor yang dihasilkan menggunakan algoritma AdaBoost.

4.1 Fitur yang digunakan

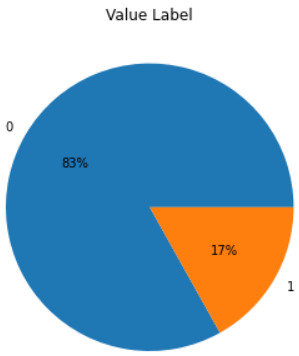
Terdapat beberapa fitur yang harus diklasifikasi pada dataset CICIDS-2017 kertika akan digunakan, fitur yang terdapat pada dataset tersebut dianalisis untuk digunakan ke dalam model AdaBoost. Dataset CICIDS-2017 memiliki 79 fitur dan setelah dilakukan *cluster* atau pengelompokan data berdasarkan kemiripan atau jarak antar data menghasilkan 76 fitur yang nantinya akan digunakan dalam proses model. Berikut adalah atribut yang ada pada datase CICIDS-2017:

Table 2 Fitur Dataset CICDS-2017 Setelah Dilakukan Cluster

No	Fitur	No	Fitur
1	Flow Duration	39	Max Packet Length
2	Total Fwd Packets	40	Packet Length Mean
3	Total Backward Packets	41	Packet Length Std
4	Total Length of Fwd Packets	42	Packet Length Variance
5	Total Length of Bwd Packets	43	FIN Flag Count
6	Fwd Packet Length Max	44	SYN Flag Count
7	Fwd Packet Length Min	45	RST Flag Count
8	Fwd Packet Length Mean	46	PSH Flag Count
9	Fwd Packet Length Std	47	ACK Flag Count
10	Bwd Packet Length Max	48	URG Flag Count
11	Bwd Packet Length Min	49	CWE Flag Count
12	Bwd Packet Length Mean	50	ECE Flag Count
13	Bwd Packet Length Std	51	Down/Up Ratio
14	Flow Bytes/s	52	Average Packet Size
15	Flow Packets/s	53	Avg Fwd Segment Size
16	Flow IAT Mean	54	Avg Bwd Segment Size
17	Flow IAT Std	55	Fwd Avg Bytes/Bulk
18	Flow IAT Max	56	Fwd Avg Packets/Bulk
19	Flow IAT Min	57	Fwd Avg Bulk Rate
20	Fwd IAT Total	58	Bwd Avg Bytes/Bulk
21	Fwd IAT Mean	59	Bwd Avg Packets/Bulk
22	Fwd IAT Std	60	Bwd Avg Bulk Rate
23	Fwd IAT Max	61	Subflow Fwd Packets
24	Fwd IAT Min	62	Subflow Fwd Bytes
25	Bwd IAT Total	63	Subflow Bwd Packets'
26	Bwd IAT Mean	64	Subflow Bwd Bytes
27	Bwd IAT Std	65	Init_Win_bytes_forward
28	Bwd IAT Max	66	Init_Win_bytes_backward
29	Bwd IAT Min	67	act_data_pkt_fwd
30	Fwd PSH Flags	68	min_seg_size_forward
31	Bwd PSH Flags	69	Active Mean
32	Fwd URG Flags	70	Active Std
33	Bwd URG Flags	71	Active Max
34	Fwd Header Length	72	Active Min
35	Bwd Header Length	73	Idle Mean
36	Fwd Packets/s	74	Idle Std
37	Bwd Packets/s	75	Idle Max
38	Min Packet Length	76	Idle Min

4.2 Value Label

Setelah dilakukan cluster fitur dataset Langkah selanjutnya adalah tahapan klasifikasi value label dataset CICIDS 2017, hasil klasifikasi value label dituangkan seperti diagram pada gambar 2 seperti berikut. Dari hasil analisis yang di tampilkan dalam bentuk diagram value diatas dihasilkan data normal dan data serangan, data serangan di defenisikan dengan kode angka 1, dan data trafik normal didefenisikan dengan angka 0. Dari kedua data tersebut, dalam dataset CICIDS 2017 diperoleh data Normal sebanyak 2096484 data atau 83% dari total data yang ada. Sedangkam untuk data serangan diperoleh sebanyak 425878 data atau 17% dari total data yang ada.



Gambar 2 Value Dan Label Dataset

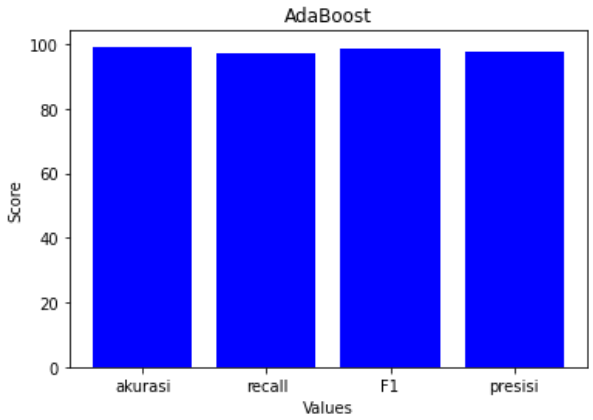
Keterangan:

Tabel 1 Data Value dan Label

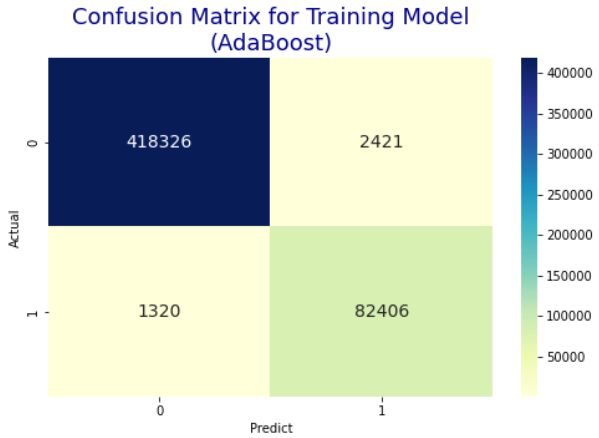
Kode	Nama	Jumlah	Persentase
0	Data Normal	2096484	83%
1	Data Serangan	425878	17%

4.3 Diagram hasil akurasi

Hasil prediksi yang dilakukan dengan menggunakan algoritma AdaBoost di peroleh beberapa nilai dari *accuracy*, *precision*, *recall*, dan F1. Berikut adalah diagram hasil yang diperoleh dari prediksi serangan DDOS:



Gambar 3 Diagram Hasil Skor



Gambar 4 Confusion Matrix

Keterangan:

Table 3 Keterangan Diagram

Jenis Skor	Hasil Skor
Accuracy	99.2 %
Precision	97.1 %
Recall	98.4 %
F1	97.7 %

5. KESIMPULAN

Hasil penelitian deteksi serangan DdoS menggunakan *machine learning* dengan algoritma *AdaBoost* diperoleh beberapa jenis skor yang terdiri dari *Accuracy* sebesar 99.2%, *Precision* sebesar 97.1%, *Recall* sebesar 98.4%, dan F1 *Score* dengan besar 97.7%. Hasil nilai akurasi akan berbeda tergantung dengan model algoritma yang diimplementasikan. Semakin tinggi akurasi sistem deteksi serangan DDoS, maka tingkat *false positive* dan *false negative* akan semakin rendah. Hal ini menunjukkan bahwa sistem tersebut mampu mendeteksi serangan DDoS dengan tepat dan tidak terlalu banyak mengalami kesalahan dalam mendeteksi serangan.

REFERENSI

Alfidzar, H., & Zen, B. P. (2022). Implementasi HoneyPy Dengan Malicious Traffic Detection System (Maltrail) Menggunakan Analisis Deskriptif Guna Untuk Mendeteksi Serangan DDOS Pada Server. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, *4*(2), 32–45. https://doi.org/10.20895/inista.v4i2.534

Amarudin, A., & Yuliansyah, A. (2018). Analisis penerapan mikrotik router sebagai user manager untuk menciptakan internet sehat. *Tam*, *9*(1), 62–66.

Ashfaq, M. F., Malik, M., Fatima, U., & Shahzad, M. K. (2022). Classification of IoT based DDoS Attack using Machine Learning Techniques. *2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM)*.

Deepa, V., Muthamil Sudar, K., & Deepalakshmi, P. (2018). Detection of DDoS attack on SDN control plane using hybrid machine learning techniques. *Proceedings of the International Conference on Smart Systems and Inventive Technology, ICSSIT 2018*, *Icssit*, 299–303. https://doi.org/10.1109/ICSSIT.2018.8748836

Firdaus, D., Munadi, R., & Purwanto, Y. (2020). DDoS Attack Detection in Software Defined Network using

Ensemble K-means++ and Random Forest. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, 164–169. https://doi.org/10.1109/ISRITI51436.2020.9315521

Hu, J. (2017). Automated detection of driver fatigue based on AdaBoost classifier with EEG signals. *Frontiers in Computational Neuroscience*, *11*(August), 1–10. https://doi.org/10.3389/fncom.2017.00072

Khuphiran, P., Leelaprute, P., Uthayopas, P., Ichikawa, K., & Watanakeesuntorn, W. (2018). Performance comparison of machine learning models for DDoS attacks detection. *2018 22nd International Computer Science and Engineering Conference, ICSEC 2018*, 1–4. https://doi.org/10.1109/ICSEC.2018.8712757

Kurniabudi, Stiawan, D., Darmawijoyo, Bin Idris, M. Y. Bin, Bamhdi, A. M., & Budiarto, R. (2020). CICIDS-2017 Dataset Feature Analysis with Information Gain for Anomaly Detection. *IEEE Access*, *8*, 132911–132921. https://doi.org/10.1109/ACCESS.2020.3009843

Shailesh Singh Panwar, Raiwani, Y. P., & Panwar, L. S. (2022). An Intrusion Detection Model for CICIDS-2017 Dataset Using Machine Learning Algorithms. *2022 International Conference on Advances in Computing, Communication and Materials (ICACCM)*.

Sumadi, F. D. S., & Aditya, C. S. K. (2021). Machine learning in openflow network: Comparative analysis of ddos detection techniques. *International Arab Journal of Information Technology*, *18*(2), 221–226. https://doi.org/10.34028/IAJIT/18/2/11

Yadahalli, S., & Nighot, M. K. (2018). Adaboost based parameterized methods for wireless sensor networks. *Proceedings of the 2017 International Conference On Smart Technology for Smart Nation, SmartTechCon 2017*, 1370–1374. https://doi.org/10.1109/SmartTechCon.2017.8358590

Yang, L., & Zhao, H. (2019). DDoS attack identification and defense using SDN based on machine learning method. *Proceedings - 2018 15th International Symposium on Pervasive Systems, Algorithms and Networks, I-SPAN 2018*, 174–178. https://doi.org/10.1109/I-SPAN.2018.00036