Rekayasa Fitur Berbasis *Machine Learning* untuk Mendeteksi Serangan DDoS

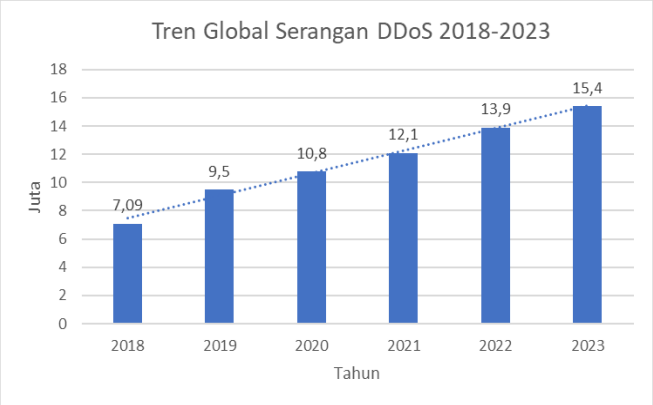
Muhammad Nur Faiz[1], Oman Somantri[2], Arif Wirawan Muhammad[3]

***Intisari*—Serangan jaringan terdistribusi yang disebut juga dengan *distributed denial of service* (DDoS) merupakan ancaman dan masalah utama keamanan internet. DDoS adalah serangan pada jaringan yang bertujuan melumpuhkan sumber daya server. Serangan ini meningkat setiap tahunnya, terlebih pada kondisi pandemi COVID-19 saat ini. Salah satu bentuk penanggulangan untuk meminimalkan dampak DDoS adalah dengan perintah deteksi sistem atau *intrusion detection system* (IDS). Teknik IDS saat ini masih banyak menggunakan metode tradisional, sehingga masih jauh dari sempurna dibandingkan dengan teknik dan alat yang digunakan penyerang, karena IDS dengan metode tradisional hanya menggunakan deteksi berbasis *signature* atau model deteksi berbasis anomali dan menyebabkan banyak kesalahan. Lalu lintas paket data jaringan memiliki sifat yang kompleks, baik dari segi ukuran maupun sumbernya. Penelitian ini memanfaatkan kemampuan jaringan saraf tiruan untuk mendeteksi serangan DDoS atau normal. Teknik klasifikasi dengan metode jaringan saraf tiruan menjadi salah satu solusi. Berdasarkan kekurangan pada IDS tradisional, penelitian ini bertujuan untuk mendeteksi serangan DDoS menggunakan teknik rekayasa fitur berbasis *feeder machine learning* untuk meningkatkan pengembangan IDS. Selain itu, penelitian ini juga bertujuan untuk menganalisis dan mendapatkan kombinasi fungsi pelatihan dan arsitektur lapisan tersembunyi jaringan saraf tiruan terbaik untuk menyelesaikan permasalahan deteksi dan klasifikasi paket DDoS dalam jaringan komputer dengan memanfaatkan *dataset* UNSW-NB15 menggunakan metode jaringan saraf tiruan, sehingga didapatkan suatu kombinasi antara fungsi pelatihan dan arsitektur jaringan tersembunyi jaringan saraf tiruan yang mampu memberikan tingkat akurasi pengenalan DDoS yang tinggi. Berdasarkan percobaan yang dilakukan dengan tiga skema dan menggunakan teknik arsitektur skema jaringan saraf tiruan dengan masukan delapan fitur, diperoleh akurasi tertinggi sebesar 98,22%. Pemilihan fitur memainkan peranan penting dalam ketepatan hasil deteksi dan kinerja pembelajaran mesin dalam masalah klasifikasi.**

***Kata Kunci*—DDoS, Seleksi Fitur, Jaringan Saraf Tiruan, *Machine Learning*.**

I. PENDAHULUAN

Dalam satu dekade terakhir, seiring dengan pengaruh dan perkembangan teknologi informasi dan komunikasi (TIK) yang semakin pesat, aktivitas sehari-hari masyarakat kini semakin banyak dilakukan dengan menggunakan internet. Akibatnya, jumlah lalu lintas jaringan meningkat dan skala infrastruktur jaringan berkembang pesat [1]. TIK sendiri telah menjadi bagian penting dalam kehidupan modern. Seiring dengan banyaknya penggunaan TIK, banyak jenis serangan terhadap jaringan telah digunakan, termasuk *denial of service* (DoS), serangan *man-in-the-middle*, *sniffer*, dan *malware* [2], [3].



Gbr. 1 Tren global serangan DDoS 2018-2023.

DoS merupakan salah satu ancaman dalam serangan infrastruktur jaringan. Salah satu varian DoS adalah distributed denial of service (DDoS). Berdasarkan Gbr. 1 [4], diperkirakan serangan DDoS akan terus meningkat setiap tahunnya. Hal ini diperparah dengan adanya pandemi COVID-19 yang mengakibatkan semakin banyak manusia beraktivitas dengan memanfaatkan teknologi internet. Terlihat pada Gbr. 1 bahwa pada tahun 2018 terdapat 7,9 juta serangan dan angka tersebut terus meningkat hingga diprediksi pada tahun 2023 mencapai 15,4 juta serangan. Dapat disimpulkan bahwa kenaikan rata-rata setiap tahunnya adalah 1,5 juta. Serangan DDoS menimbulkan ancaman bagi pengguna internet dan semua infrastruktur yang ada di dalamnya, termasuk *bandwidth*, sumber daya server, integritas data, ketersediaan data, dan kerahasiaan data yang tersimpan di server [5]. Hingga saat ini, serangan DDoS masih termasuk sebagai ancaman utama keamanan siber. Deteksi dini memainkan peranan mendasar dalam mencegah dampak fatal serangan DDoS pada sumber daya server [6]. Salah satu tindakan dasar yang dilakukan untuk mencegah serangan DDoS adalah dengan memasang *intrusion detection system* (IDS) di server untuk memantau aliran paket data yang masuk ke jaringan internal atau sebaliknya [7]. Sistem deteksi pada IDS hanya memantau dan memberikan *tag* (penanda) terhadap aktivitas jaringan yang mencurigakan dan langsung dilaporkan sebagai *alert* (peringatan). Hal ini mengakibatkan adanya volume *alert* yang terlalu besar karena tingginya tingkat rata-rata kesalahan pengenalan paket data normal sebagai paket DDoS atau sebaliknya, yang disebabkan oleh lalu lintas data jaringan yang bersifat nonstasioner.

[1,2] *Jurusan Teknik Informatika, Politeknik Negeri Cilacap, 53212 Cilacap, INDONESIA (tlp: 0282-533329; fax: 0282-537992; email: [1]faiz@pnc.ac.id, [2]oman.somantri@pnc.ac.id)*

[3] *Teknik Informatika, Fakultas Informatika, IT Telkom Purwokerto, 53147 Purwokerto, INDONESIA (telp: 0281-641629; email: [3]arif@ittelkom-pwt.ac.id)*

[Diterima: 18 November 2021, Revisi: 5 Juli 2022]

Deteksi intrusi umumnya terdiri atas dua pendekatan. Pendekatan pertama adalah deteksi berbasis *signature* karena *alert* dihasilkan berdasarkan *signature* serangan yang spesifik [8]. Dalam proses deteksi dari pendekatan berbasis *signature*, IDS tidak dapat mendeteksi serangan yang belum dikenal karena basis data (*database*) *signature* yang telah kedaluwarsa atau karena *signature* memang belum tersedia. Pendekatan kedua adalah deteksi berbasis anomali. Pada deteksi ini, perlu diciptakan suatu profil perilaku khas dalam taraf tertentu dari fitur aktivitas jaringan. Profil ini kemudian dijadikan sebagai dasar untuk mendefinisikan aktivitas jaringan normal [9]. Jika ada aktivitas jaringan menyimpang terlalu jauh dari profil, *alert* akan terbentuk. IDS berbasis anomali memiliki keunggulan yaitu dapat mendeteksi teknik serangan baru [10]. Di sisi lain, IDS berbasis anomali lebih kompleks jika dibandingkan dengan IDS berbasis *signature* [11]. Dengan demikian, model pendeteksian secara logika akan banyak menimbulkan *false positive flags*, karena aliran paket data jaringan komputer memiliki sifat dinamis, baik dari segi ukuran, sumber, protokol, maupun isi datanya [12]. Di sisi lain, IDS berbasis *signature* dan IDS berbasis anomali memiliki dua kelemahan utama. Kelemahan pertama terjadi ketika IDS mendeteksi serangan yang diawali dengan protokol SYN, misalnya *SYN flood*, karena protokol SYN merupakan protokol yang legal dan mutlak digunakan untuk memulai komunikasi antara dua komputer/perangkat dalam suatu jaringan [13]. Oleh karena itu, IDS biasa sulit menghasilkan *alert* terhadap serangan yang dimulai dengan artefak protokol SYN [14]. Kelemahan IDS berikutnya terutama disebabkan oleh defisit protokol TCP/IP yang memudahkan penyerang memulai serangan DDoS, misalnya dengan menggunakan perintah *ping* yang tersedia secara *default* di seluruh sistem operasi atau menggunakan alat khusus seperti HOIC, LOIC, XOIC, dan GoldenEye [15]. Berdasarkan kelemahan-kelemahan yang ada pada IDS biasa, penelitian ini bertujuan mendeteksi serangan DDoS dengan memanfaatkan teknik *machine learning*, sehingga dapat menjadi perbaikan dalam pengembangan perangkat IDS. Penelitian ini memanfaatkan *dataset* serangan DDoS yang bersumber dari UNSW-NB15 (University of New South Wales) untuk diproses lebih lanjut dengan menerapkan metode jaringan saraf tiruan untuk menghasilkan model *machine learning* pendeteksi DDoS.

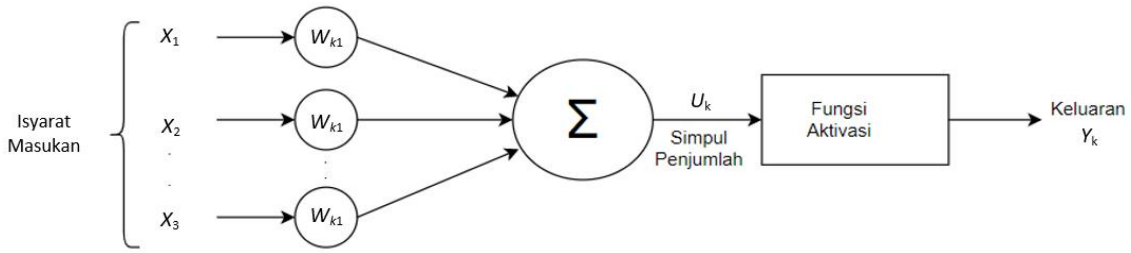
Beberapa penelitian sebelumnya terkait rekayasa fitur dalam deteksi DDoS menganalisis *dataset* UNSW-NB15 dengan menemukan relevansi fitur menggunakan jaringan saraf tiruan [16]. Penelitian tersebut mengategorikan fitur menjadi lima kelompok berdasarkan jenisnya, seperti fitur berbasis aliran, berbasis konten, berbasis waktu, esensial, dan tambahan. Dari kelompok-kelompok ini, 31 kemungkinan kombinasi fitur dievaluasi dan didiskusikan. Akurasi tertinggi dalam penelitian ini, yaitu 93%, diperoleh dengan menggunakan 39 fitur dari kelompok yang dikategorikan. Selain itu, dalam penelitian tersebut terdapat kombinasi dari 23 fitur yang dipilih menggunakan *meta-estimator* bernama SelectFromModel yang memilih fitur berdasarkan skornya. Sebanyak 23 fitur yang dipilih menghasilkan akurasi yang lebih tinggi, yaitu 97%, dibandingkan dengan 39 fitur yang sudah dijelaskan. Penelitian selanjutnya melaporkan bahwa jumlah fitur di KDD-99 lebih sedikit dari UNSW-NB15 dan kelas serangan pertama di KDD-99 (DoS) memiliki jumlah fitur paling banyak [12]. Akurasi pendeteksian serangan tersebut juga yang tertinggi, yaitu 99,40%. Hanya ada dua belas fitur untuk kelas serangan DoS di UNSW-NB15, sedangkan untuk ACC dilaporkan sebesar 86,57%. Jumlah fitur yang dipilih untuk kelas serangan KDD-99 lebih sedikit daripada UNSW-NB15. Rata-rata 25,2 fitur dipilih untuk serangan di KDD-99, sedangkan untuk serangan di UNSW-NB15, rata-rata 19,1 fitur yang dipilih. Penting untuk diperhatikan bahwa jumlah fitur terpilih terendah adalah untuk serangan Generic di UNSW-NB15, yaitu sepuluh fitur. Penelitian terkait rekayasa fitur selanjutnya mengeksplorasi penerapan algoritme XGBoost untuk pemilihan fitur dalam hubungannya dengan beberapa teknik *machine learning*, termasuk *artificial neural network* (ANN), *k-nearest neighbour* (*k*-NN), *decision tree* (DT), *logistic regression* (LR), dan *support vector machine* (SVM) untuk mengimplementasikan IDS yang akurat [7]. Metode pemilihan atribut berbasis XGBoost diterapkan pada UNSW-NB15 dan sebagai hasilnya, sembilan belas fitur optimal dipilih. Hasil dari penelitian ini juga mempertimbangkan konfigurasi klasifikasi biner dan *multi-class*. Hasilnya menunjukkan bahwa metode pemilihan fitur berbasis XGBoost memungkinkan metode seperti DT untuk meningkatkan akurasi pengujiannya dari 88,13% menjadi 90,85% untuk skema klasifikasi biner. Penelitian lainnya bertujuan untuk menggabungkan seluruh *dataset* UNSW-NB15 menjadi satu *file* sehingga dapat menguji model satu kali, bukan menguji model secara terpisah untuk setiap *file* [17]. Kemudian, digunakan *dataset* dari serangan sebagai label baru sehingga akan mengembangkan *dataset* berlabel *multi-class*. Penelitian ini juga menyelidiki kinerja pembelajaran mendalam dengan *dataset* yang ditingkatkan dalam dua kategori klasifikasi (biner dan *multi-class*). Model yang diusulkan menghasilkan akurasi 99,59% dalam klasifikasi *multi-class* dan 99,26% dalam klasifikasi biner. Penelitian ini berfokus pada pengombinasian fitur terpilih dengan teknik *information gain* yang digunakan sebagai masukan klasifikasi jaringan saraf tiruan untuk meningkatkan nilai akurasi pendeteksi DDoS.

II. PENDEKATAN DETEKSI

Pendekatan deteksi serangan DDoS yang diterapkan pada penelitian ini dibagi menjadi beberapa tahap yang dijelaskan sebagai berikut.

A. Dataset

Langkah pertama adalah mendapatkan *dataset* serangan DDoS UNSW-NB15 yang diterbitkan oleh University of New South Wales. *Dataset* UNSW-NB15 merupakan *dataset* serangan yang berisi catatan aliran paket serangan dan paket normal berupa *file tcpdump*, yang merekam aliran data selama 31 jam [12], [18]. Alur paket serangan disimulasikan secara sintetik menggunakan perangkat lunak Ixia, meniru serangan dengan *footprinting* rendah kecepatan tinggi. Terdapat sembilan jenis serangan yang tercakup dalam *dataset* UNSW-NB15, seperti disajikan pada Tabel I. Pengelompokan kategori fitur *dataset* UNSW-NB15 dilakukan secara sistematis, yaitu fitur *flow*, fitur dasar, fitur isi paket data, fitur waktu, dan fitur



Gbr. 2 Komponen jaringan saraf tiruan.

TABEL I
FITUR DDOS PADA UNSW-NB15

Nomor Fitur	UNSW-NB15	Keterangan
1	srcip	Sumber alamat IP
2	sport	Sumber nomor *port*
3	dstip	Tujuan alamat IP
4	dsport	Tujuan nomor *port*
5	proto	Protokol transaksi
6	state	Indikasi ke *state* bagian dan protokol dependennya
7	dur	Rekaman total durasi
8	sbytes	Transaksi setiap *byte* dari sumber ke tujuan
9	dbytes	Transaksi setiap *byte* dari tujuan ke sumber
10	sttl	Nilai saat waktu berjalan dari sumber ke tujuan
11	dttl	Nilai saat waktu berjalan dari tujuan ke sumber
12	sloss	Paket sumber yang dikirimkan ulang atau dijatuhkan
13	dloss	Paket tujuan yang dikirimkan ulang atau dijatuhkan
14	service	http, dns, ssh, pop3, ftp, smtp
15	stime	Rekam waktu mulai
16	Classification	0 untuk rekam jejak normal dan 1 untuk rekam jejak penyerang

tambahan. Pada dasarnya, motivasi pembentukan *dataset* UNSW-NB15 adalah untuk memperbaiki permasalahan kekurangan pada *dataset* KDD CUP 99 dan NSL-KDD [19], [20].

B. Seleksi Fitur

Pada penelitian ini, jenis rekaman yang dianalisis dikhususkan untuk kelompok rekaman DDoS seperti yang disajikan pada Tabel I [6]. *Dataset attack record* DDoS UNSW-NB15 memiliki fitur seperti pada Tabel I. Keenam belas fitur tersebut kemudian dipilih menggunakan teknik *information gain* dengan tujuan mengurangi waktu komputasi dan mendapatkan model *machine learning* dengan akurasi tinggi. *Information gain* adalah jumlah *mutual information* yang diperoleh dari kombinasi variabel observasional dan merupakan divergensi dari teori Kullback-Leibler [21]. Dalam hal *machine learning*, *information gain* berguna untuk menyeleksi beberapa fitur penting berdasarkan teori yang mengukur nilai informasi yang dimiliki oleh sebuah fitur yang berhubungan dengan fitur lainnya. Untuk fitur "a", *information gain* adalah jumlah entropi yang dikandung oleh "a" dibandingkan dengan fitur "c" dari semua fitur yang tersedia [22]. Fitur penting ditunjukkan dengan nilai maksimum entropi yang dimiliki fitur tersebut. Persamaan *information gain* disajikan pada (1).

$$info\ gain_a = H(C) - H\left(\frac{C}{a}\right)$$
$$H(c) = -\sum_{c\epsilon C} P(c) * \log_2 P(c) \quad (1)$$
$$H\left(\frac{C}{a}\right) = -\sum_{c\epsilon A} P(c) * \sum_{c\epsilon C} p\left(\frac{C}{a}\right) * \log_2 P\left(\frac{C}{a}\right)$$

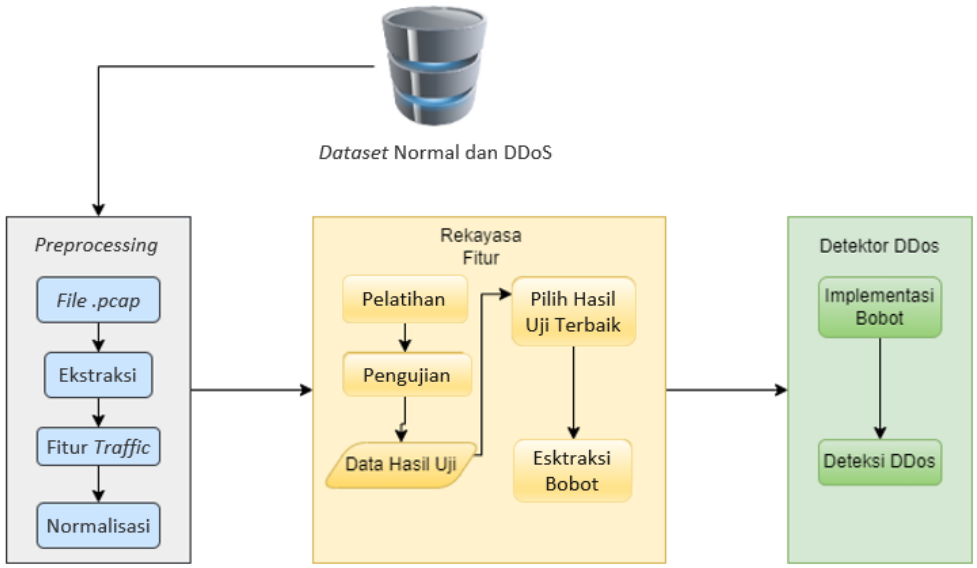
dengan *H* merupakan entropi dan *P* adalah probabilitas.

C. Jaringan Saraf Tiruan

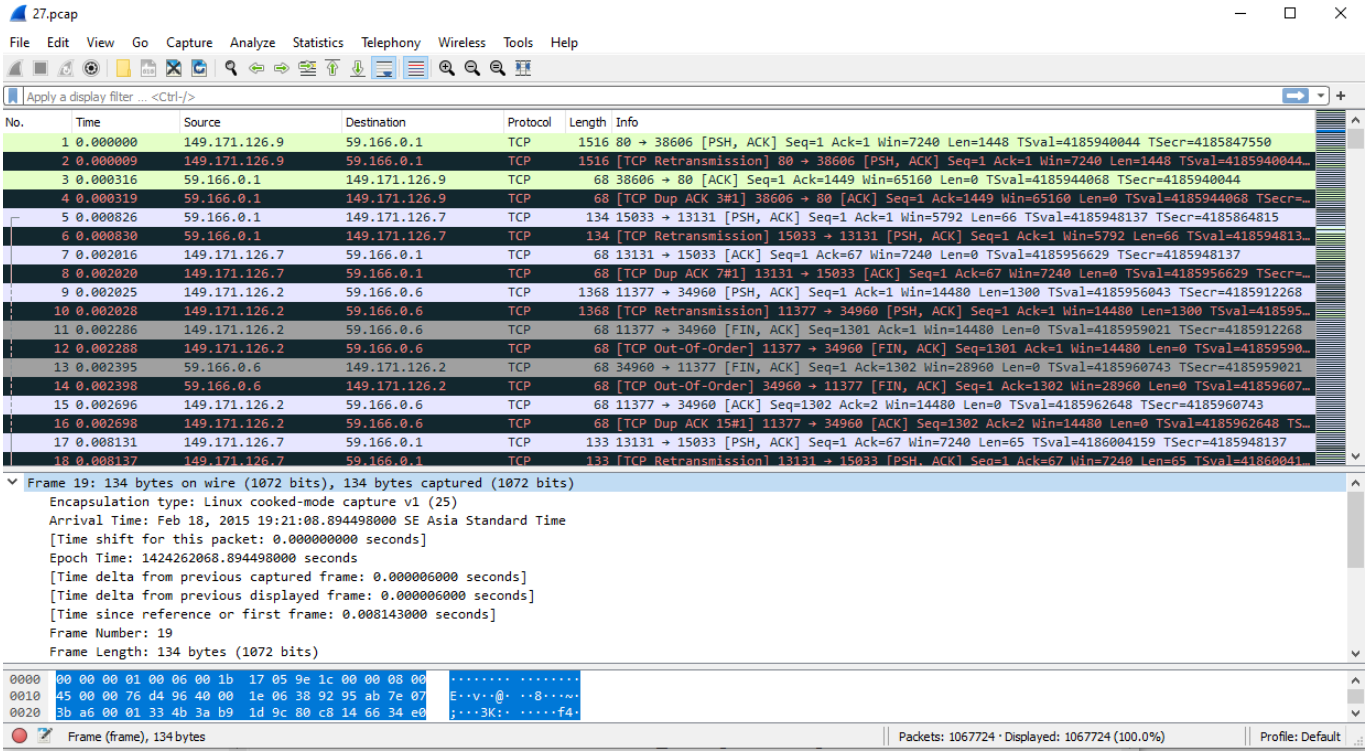
Jaringan saraf tiruan adalah paradigma pemrosesan informasi yang terinspirasi oleh sistem sel saraf biologi, sama seperti otak yang memproses suatu informasi. Pada jaringan otak manusia terdapat sel saraf (neuron) yang memiliki tiga komponen penyusun yang saling bekerja sama untuk mengolah sinyal-sinyal informasi. Tiga komponen tersebut adalah dendrit (masukan), badan sel (pengolah masukan), dan akson (keluaran) [23].

Gbr. 2 mengilustrasikan komponen jaringan saraf tiruan. *X* adalah masukan yang akan dikirim ke neuron dengan bobot kedatangan tertentu. Masukan ini akan diproses oleh suatu fungsi perambatan yang menjumlahkan nilai semua bobot yang datang yang disimbolkan dengan W_{k1}. Hasil penjumlahan tersebut dibandingkan dengan suatu nilai ambang (*threshold*) tertentu melalui fungsi aktivasi setiap neuron. Apabila masukan melewati nilai ambang tertentu, neuron akan diaktifkan. Jika tidak, neuron dinonaktifkan. Apabila neuron diaktifkan, neuron mengirimkan keluaran melalui bobot-bobot keluarannya ke semua neuron yang berhubungan dengannya yang disimbolkan dengan Y_k.

Lapisan tersembunyi merupakan tiruan dari sel-sel saraf konektor pada jaringan saraf biologis. Lapisan tersembunyi berfungsi meningkatkan kemampuan jaringan saraf tiruan dalam menyelesaikan suatu masalah. Konsekuensi dari adanya lapisan ini adalah pelatihan menjadi makin sulit atau lama. Jumlah lapisan tersembunyi yang makin banyak akan meningkatkan kemampuan jaringan saraf tiruan dalam menyelesaikan masalah yang kompleks. Namun, di sisi lain, hal ini akan memperlama proses pembelajaran dan menurunkan kinerja jaringan. Secara teori, penggunaan satu lapisan tersembunyi pada jaringan saraf tiruan sudah cukup untuk menyelesaikan sebuah kasus prediksi [23]. Kolmogorov



Gbr. 3 Kerangka kerja penelitian.



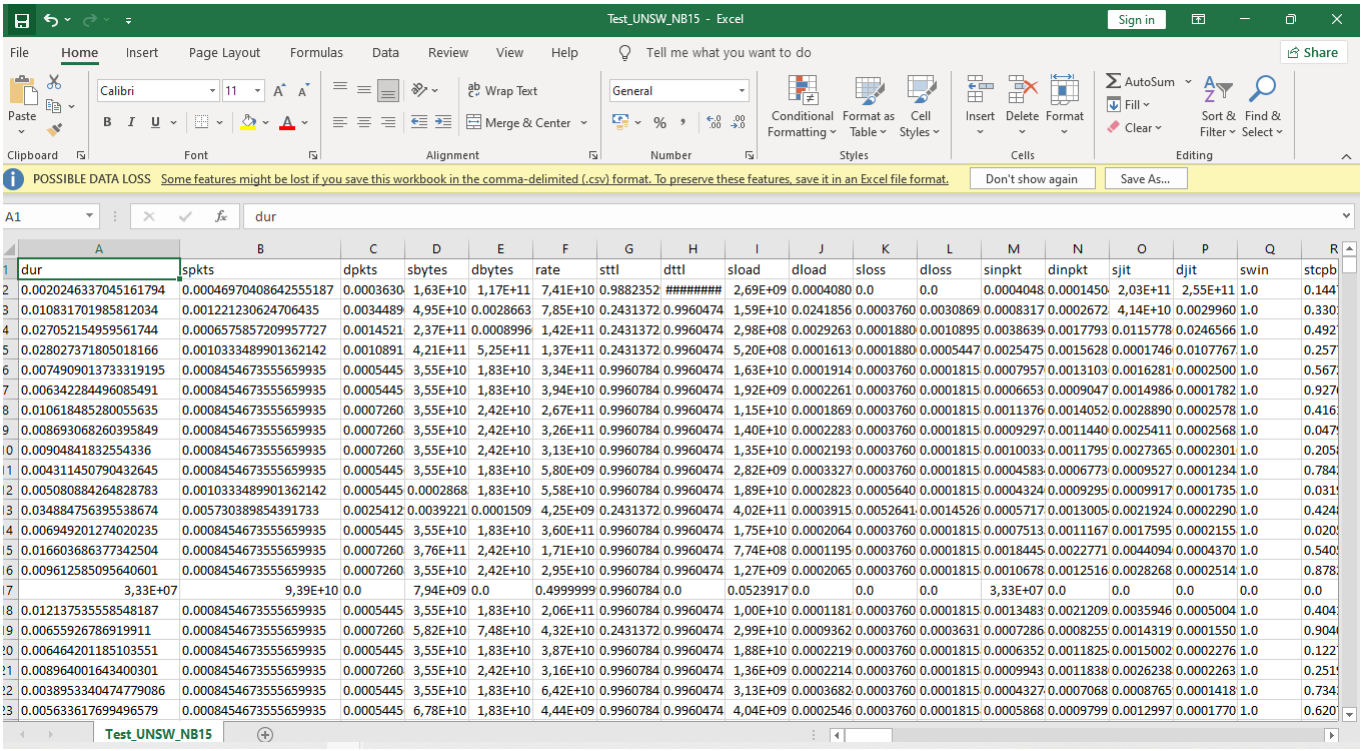
Gbr. 4 *Dataset* berupa *.pcap* dari UNSW-NB15.

menyebutkan bahwa jumlah lapisan tersembunyi terbaik untuk menyelesaikan suatu permasalahan dengan jaringan saraf tiruan adalah 2*n* + 1, dengan *n* adalah jumlah neuron masukan [24].

D. Matriks Kinerja

Pada penelitian ini digunakan beberapa parameter yang berfungsi untuk memudahkan analisis kinerja pengenalan. Indikator yang digunakan adalah *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN) [25]. TP adalah pengenalan paket data DDoS yang diidentifikasi oleh jaringan saraf tiruan sebagai suatu paket DDoS, sedangkan TN adalah pengenalan paket data normal yang diidentifikasi oleh jaringan saraf tiruan sebagai suatu paket normal. FP adalah pengenalan paket data normal yang diidentifikasi oleh jaringan saraf tiruan sebagai paket DDoS dan FN adalah pengenalan paket data DDoS yang diidentifikasi oleh jaringan saraf tiruan sebagai paket normal. Dari indikator-indikator tersebut, dapat dibentuk persamaan yang menyatakan akurasi, yang disajikan pada (2).

$$Akurasi = \frac{TN+TP}{FP+FN+TP+TN} \quad (2)$$



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	dur	spkts	dpkts	sbytes	dbytes	rate	sttl	dttl	sload	dload	sloss	dloss	sinpkt	dinpkt	sjit	djit	swin	stcpb
2	0.00202463370451617 94	0.00046970408642555187	0.0003630	1,63E+10	1,17E+11	7,41E+10	0.9882352	#########	2,69E+09	0.0004080	0.0	0.0	0.0004048	0.0001450	2,03E+11	2,55E+11	1.0	0.144
3	0.010831701985812034	0.001221230624706435	0.0034489	4,95E+10	0.0028663	7,85E+10	0.2431372	0.9960474	1,59E+10	0.0241856	0.0003760	0.0030869	0.0008317	0.0002672	4,14E+10	0.0029960	1.0	0.330
4	0.027052154959561744	0.00065758572099577 27	0.0014521	2,37E+11	0.0008996	1,42E+11	0.2431372	0.9960474	2,98E+08	0.0029263	0.0001880	0.0010895	0.0038639	0.0017793	0.0115778	0.0246566	1.0	0.492
5	0.028027371805018166	0.0010333489901362142	0.0010891	4,21E+11	5,25E+11	1,37E+11	0.2431372	0.9960474	5,20E+08	0.0001613	0.0001880	0.0005447	0.0025475	0.0015628	0.0001746	0.0107767	1.0	0.257
6	0.007490901373331919 5	0.0008454673555659935	0.0005445	3,55E+10	1,83E+10	3,34E+11	0.9960784	0.9960474	1,63E+10	0.0001914	0.0003760	0.0001815	0.0007957	0.0013103	0.0016281	0.0002500	1.0	0.567
7	0.006342284496085491	0.0008454673555659935	0.0005445	3,55E+10	1,83E+10	3,94E+10	0.9960784	0.9960474	1,92E+09	0.0002261	0.0003760	0.0001815	0.0006653	0.0009047	0.0014986	0.0001782	1.0	0.927
8	0.010618485280055635	0.0008454673555659935	0.0007260	3,55E+10	2,42E+10	2,67E+11	0.9960784	0.9960474	1,15E+10	0.0001869	0.0003760	0.0001815	0.0011376	0.0014052	0.0028890	0.0002578	1.0	0.416
9	0.008693068260395849	0.0008454673555659935	0.0007260	3,55E+10	2,42E+10	3,26E+11	0.9960784	0.9960474	1,40E+10	0.0002283	0.0003760	0.0001815	0.0009297	0.0011440	0.0025411	0.0002568	1.0	0.047
10	0.00904841832554336	0.0008454673555659935	0.0007260	3,55E+10	2,42E+10	3,13E+10	0.9960784	0.9960474	1,35E+10	0.0002193	0.0003760	0.0001815	0.0010033	0.0011795	0.0027365	0.0002301	1.0	0.205
11	0.004311450790432645	0.0008454673555659935	0.0005445	3,55E+10	1,83E+10	5,80E+09	0.9960784	0.9960474	2,82E+09	0.0003327	0.0003760	0.0001815	0.0004583	0.0006773	0.0009527	0.0001234	1.0	0.784
12	0.005080884264828783	0.0010333489901362142	0.0005445	0.0002868	1,83E+10	5,58E+10	0.9960784	0.9960474	1,89E+10	0.0002823	0.0005640	0.0001815	0.0004324	0.0009295	0.0009917	0.0001735	1.0	0.031
13	0.034884756395538674	0.005730389854391733	0.0025412	0.0039221	0.0001509	4,25E+09	0.2431372	0.9960474	4,02E+11	0.0003915	0.0052641	0.0014526	0.0005717	0.0013005	0.0021924	0.0002290	1.0	0.424
14	0.006949201274020235	0.0008454673555659935	0.0005445	3,55E+10	1,83E+10	3,60E+11	0.9960784	0.9960474	1,75E+10	0.0002064	0.0003760	0.0001815	0.0007513	0.0011167	0.0017595	0.0002155	1.0	0.020
15	0.016603686377342504	0.0008454673555659935	0.0007260	3,76E+11	2,42E+10	1,71E+10	0.9960784	0.9960474	7,74E+08	0.0001195	0.0003760	0.0001815	0.0018445	0.0022771	0.0044094	0.0004370	1.0	0.540
16	0.00961258095640601	0.0008454673555659935	0.0007260	3,55E+10	2,42E+10	2,95E+10	0.9960784	0.9960474	1,27E+09	0.0002065	0.0003760	0.0001815	0.0010678	0.0012516	0.0028268	0.0002514	1.0	0.878
17	3,33E+07	9,39E+10	0.0	7,94E+09	0.0	0.4999999	0.9960784	0.0	0.0523917	0.0	0.0	0.0	3,33E+07	0.0	0.0	0.0	0.0	0.0
18	0.012137535558548187	0.0008454673555659935	0.0005445	3,55E+10	1,83E+10	2,06E+11	0.9960784	0.9960474	1,00E+10	0.0001181	0.0003760	0.0001815	0.0013483	0.0021209	0.0035946	0.0005004	1.0	0.404
19	0.00655926786919911	0.0008454673555659935	0.0007260	5,82E+10	7,48E+10	4,32E+10	0.2431372	0.9960474	2,99E+10	0.0009362	0.0003760	0.0003631	0.0007286	0.0008255	0.0014319	0.0001550	1.0	0.904
20	0.006464201185103551	0.0008454673555659935	0.0005445	3,55E+10	1,83E+10	3,87E+10	0.9960784	0.9960474	1,88E+10	0.0002219	0.0003760	0.0001815	0.0006352	0.0011825	0.0015002	0.0002276	1.0	0.122
21	0.008964001643400301	0.0008454673555659935	0.0007260	3,55E+10	2,42E+10	3,16E+10	0.9960784	0.9960474	1,36E+09	0.0002214	0.0003760	0.0001815	0.0009943	0.0011838	0.0026238	0.0002263	1.0	0.251
22	0.0038953340474779086	0.0008454673555659935	0.0005445	3,55E+10	1,83E+10	6,42E+10	0.9960784	0.9960474	3,13E+09	0.0003682	0.0003760	0.0001815	0.0004327	0.0007068	0.0008765	0.0001418	1.0	0.734
23	0.005633617699496579	0.0008454673555659935	0.0005445	6,78E+10	1,83E+10	4,44E+09	0.9960784	0.9960474	4,04E+09	0.0002546	0.0003760	0.0001815	0.0005868	0.0009799	0.0012997	0.0001770	1.0	0.620

Gbr. 5 Ekstraksi *dataset*.

Akurasi adalah rasio antara pengenalan paket DDoS yang ditambah dengan pengenalan paket normal terhadap keseluruhan paket data. *TP* merupakan tingkat contoh yang diidentifikasi dengan benar sebagai serangan; *TN* adalah tingkat lalu lintas yang sah yang diklasifikasikan sebagai sah; *FP*, kadang-kadang disebut sebagai kesalahan tipe I, adalah tingkat lalu lintas yang sah yang diklasifikasikan sebagai serangan; sedangkan *FN*, kadang-kadang disebut sebagai kesalahan tipe II, adalah tingkat lalu lintas yang sah yang diklasifikasikan sebagai intrusi.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Presisi = \frac{TP}{TP+FN} \quad (4)$$

$$F1-Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \quad (5)$$

Recall/*sensitivity* menyatakan aktual positif dikategorikan dengan benar sebagai kelas positif, presisi merupakan ukuran estimasi probabilitas prediksi positif yang benar, sedangkan *F-score*/*F-measure* merupakan perbandingan rata-rata presisi dan *recall* yang dibobotkan.

E. Kerangka Kerja Penelitian

Langkah yang digunakan dalam penelitian ini ditunjukkan pada Gbr. 3, yang dijelaskan sebagai berikut. Pertama, dilakukan pengambilan *dataset* paket data serangan DDoS dan aliran paket data normal UNSW-NB15 dalam bentuk *.pcap* [18]. Lalu dilakukan transformasi *dataset .pcap* menjadi format *.csv* untuk diekstraksi isinya. Data hasil ekstraksi selanjutnya dikuantifikasi untuk mendapatkan fitur lalu lintas jaringan. Selanjutnya, dilaksanakan proses normalisasi data hasil kuantifikasi dengan cara membagi setiap data hasil kuantifikasi dengan nilai maksimumnya, sehingga didapatkan nilai data relatif dengan nilai maksimal sebesar 1. Langkah berikutnya adalah melakukan pelatihan data jaringan saraf dari hasil normalisasi dengan beberapa kriteria pelatihan. Setelah itu, dilakukan ekstraksi data asli *log* server dalam bentuk entitas-properti. Rekayasa fitur (*feature engineering*) lalu dilakukan terhadap data asli *log* server, yaitu dengan *compounding feature*, *splitting feature*, *merging feature*, serta *one-hot encoding*. Hasil rekayasa fitur ini dimasukkan ke *machine learning* untuk proses deteksi aliran paket data normal dan serangan DDoS. Terakhir, dilakukan analisis terhadap kinerja klasifikasi *machine learning* dalam mendeteksi aliran paket data normal dan serangan DDoS berdasarkan metrik akurasi, presisi, *recall*, dan *F1-score*.

III. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan menggunakan perangkat lunak MATLAB 2019b yang berjalan pada platform sistem operasi Windows 10 64-bit. *Dataset* UNSW-NB15 ini cukup banyak pada basis datanya, sehingga hanya diambil seri *27.pcap* atau sesudahnya. Eksperimen ini dilatih, dievaluasi, dan diuji pada Microsoft Excel dan MATLAB.

Gbr. 4 menunjukkan *file .pcap* atau data mentah dari UNSW-NB15. Data tersebut merupakan data yang siap diolah dan dikonversi ke *.csv* agar lebih mudah dianalisis. Terdapat 1.067.724 *dataset* pada *file 27.pcap*. Gbr. 5 menunjukkan hasil konversi dari *.pcap* ke *.csv*. Data tersebut nantinya akan diolah dan diuji. Pada data terebut terdapat 49 fitur, kemudian disederhanakan dengan mencari indeks bias per blok sesuai fitur yang dipilih.

TABEL II
SKEMA ARSITEKTUR JARINGAN SARAF TIRUAN

Nomor Skema	Neuron Masukan	Total Neuron Lapisan Tersembunyi	Neuron Keluaran
1	7	15	2 (Normal dan DDoS)
2	8	17	2 (Normal dan DDoS)
3	15	31	2 (Normal dan DDoS)

TABEL III
HASIL PERCOBAAN DENGAN SKEMA 1

Percobaan	Akurasi	Presisi	*Recall*	*F1-Score*
1	98,24%	98,53%	97,52%	98,03%
2	97,71%	96,73%	96,68%	96,71%
3	97,34%	96,95%	96,22%	96,58%

TABEL IV
HASIL PERCOBAAN DENGAN SKEMA 2

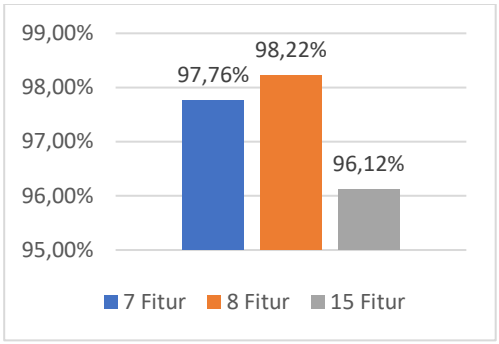
Percobaan	Akurasi	Presisi	*Recall*	*F1-Score*
1	97,79%	98,12%	99,72%	98,92%
2	99,73%	98,89%	99,94%	99,41%
3	97,15%	97,27%	99,12%	98,19%

Dalam penelitian ini, tiga skema fitur digunakan sebagai masukan dari pengklasifikasi jaringan saraf tiruan untuk mengetahui efektivitas pelatihan dan akurasi klasifikasi yang dihasilkan dari proses pemilihan fitur. Berdasarkan skema masukan seleksi fitur, terbentuk tiga skema arsitektur jaringan saraf tiruan yang berbeda, yang didasarkan pada teori bahwa penggunaan satu lapisan tersembunyi pada jaringan saraf tiruan sudah cukup untuk menyelesaikan sebuah kasus prediksi [26]. Skema arsitektur jaringan saraf tiruan tersebut tersaji pada Tabel II. Di sisi lain, Kolmogorov menyebutkan bahwa jumlah lapisan tersembunyi terbaik untuk menyelesaikan suatu permasalahan dengan jaringan saraf tiruan adalah $2n + 1$, dengan n adalah jumlah neuron masukan. Berdasarkan teori yang dipaparkan oleh Kolmogorov tersebut, pada penelitian ini dibentuk variasi arsitektur jaringan saraf tiruan untuk mencari akurasi tertinggi dalam menyelesaikan permasalahan deteksi paket jaringan DDoS [24], [27].

Pada penelitian ini, dilakukan tiga kali percobaan dengan skema dan *dataset* dari UNSW-NB15 untuk mencari tingkat akurasi tertinggi. Tingkat akurasi ini didasarkan pada masukan fitur. Proses rekayasa fitur dilakukan dengan memilih fitur kemudian mengombinasikannya sesuai dengan Tabel II. Hasil percobaan pertama ditunjukkan pada Tabel III. Terlihat bahwa akurasi tertinggi dari percobaan pertama adalah 98,24%. Percobaan pertama juga memiliki nilai presisi, *recall*, dan *F1-score* lebih tinggi dibandingkan yang lain. Skema 1 ini merupakan skema dengan pemilihan tujuh fitur.

Tabel IV memperlihatkan bahwa percobaan kedua mendapatkan akurasi tertinggi dengan nilai 99,73%. Percobaan kedua juga memiliki nilai presisi, *recall*, dan *F1-score* lebih tinggi dibandingkan yang lain. Skema ini adalah skema dengan pemilihan sembilan fitur.

Tabel V menunjukkan bahwa akurasi pada percobaan ketiga adalah 97,13%. Percobaan ketiga juga memiliki nilai presisi, *recall*, dan *F1-score* lebih tinggi dibandingkan yang lain. Skema 3 merupakan skema dengan pemilihan lima belas fitur.



Gbr. 6 Grafik tingkat akurasi percobaan.

TABEL V
HASIL PERCOBAAN DENGAN SKEMA 3

Percobaan	Akurasi	Presisi	Recall	*F1-Score*
1	95,02%	93,20%	96,74%	97,10%
2	97,13%	96,52%	99,29%	97,93%
3	96,22%	95,55%	95,01%	96,00%

TABEL VI
KESIMPULAN HASIL PERCOBAAN DENGAN TIGA SKEMA

Jumlah Fitur yang Terseleksi	Akurasi	Presisi	*Recall*	*F1-Score*
Jumlah fitur ada 7 dengan nomor fitur: 2, 6, 9, 10, 12, 1, 3	97,76%	97,40%	96,81%	97,10%
Jumlah fitur ada 8 dengan nomor fitur: 2, 6, 9, 10, 12, 15, 1, 3	98,22%	98,09%	99,59%	97,93%
Jumlah fitur ada 15	96,12%	95,02%	97,01%	96,00%

Tabel VI menyajikan kesimpulan hasil eksperimen yang dilakukan. Dapat dilihat bahwa dengan tujuh fitur yang diseleksi, dapat diperoleh tingkat akurasi 97,76%, sedangkan dengan memilih hampir semua fitur (lima belas), justru tingkat akurasi lebih kecil. Eksperimen selanjutnya dengan kombinasi delapan fitur memperoleh 98,22%. Hasil eksperimen dengan delapan fitur ini menunjukkan akurasi tertinggi. Grafik akurasi ditunjukkan pada Gbr. 6.

IV. KESIMPULAN

Berdasarkan hasil eksperimen yang telah dilakukan, ditemukan bahwa seleksi fitur berperan penting dalam menentukan akurasi hasil deteksi dan efisiensi pelatihan *machine learning* dalam masalah klasifikasi. Pada penelitian ini, kombinasi delapan fitur utama dari *dataset* yang digunakan sebagai masukan klasifikasi jaringan saraf tiruan, yaitu fitur nomor 2, 6, 9, 10, 12, 15, 1, dan 3, menghasilkan nilai akurasi tertinggi sebesar 98,22%, dibandingkan kedua fitur lainnya. Skema kombinasi delapan fitur tersebut juga menghasilkan model jaringan saraf tiruan yang memiliki efisiensi pelatihan terbaik. Pada akhirnya, dapat disimpulkan bahwa untuk menutupi kekurangan IDS biasa dalam menyelesaikan masalah pendeteksian serangan DDoS, berdasarkan *dataset* UNSW-NB15 dan jaringan saraf tiruan, dibutuhkan delapan fitur terpilih dari enam belas fitur yang tersedia.

Penelitian selanjutnya diharapkan dapat mengombinasikan beberapa fitur dan algoritme yang lain agar dapat meningkatkan akurasi dalam deteksi serangan DDoS.

KONFLIK KEPENTINGAN

Selama penulisan dan jalannya penelitian ini, penulis tidak memiliki konflik dengan berbagai pihak mana pun. Setiap informasi yang disampaikan merupakan hasil asli sebagaimana yang didapatkan saat melakukan penelitian dan tidak dipengaruhi oleh opini atau kepentingan pribadi.

KONTRIBUSI PENULIS

Penelitian ini merupakan penelitian kolaboratif yang dilakukan oleh tiga penulis. Berikut ini merupakan pembagian kontribusi masing-masing penulis yang terlibat dalam penelitian ini. Konseptualisasi, Muhammad Nur Faiz dan Oman Somantri; penulisan penyusunan draf asli dan administrasi, Muhammad Nur Faiz, Oman Somantri, dan Arif Wirawan Muhammad; perangkat lunak dan pengumpulan data, Muhammad Nur Faiz; pengawasan dan validasi, Arif Wirawan Muhammad.

UCAPAN TERIMA KASIH

Penelitian ini dapat terlaksana atas dukungan berbagai pihak, terutama Pusat Penelitian dan Pengabdian kepada Masyarakat Politeknik Negeri Cilacap.

REFERENSI

[1] K. Kurniabudi, A. Harris, dan A. Rahim, "Seleksi Fitur dengan Information Gain untuk Meningkatkan Deteksi Serangan DDoS Menggunakan Random Forest," *Techno.COM*, Vol. 19, No. 1, hal. 56–66, Feb. 2020.

[2] S. Haider, dkk., "A Deep CNN Ensemble Framework for Efficient DDoS Attack Detection in Software Defined Networks," *IEEE Access*, Vol. 8, hal. 53972–53983, Feb. 2020.

[3] H. Parmar dan A. Gosai, "Analysis and Study of Network Security at Transport Layer," *Int. J. Comput. Appl.*, Vol. 121, No. 13, hal. 35–40, Jul. 2015.

[4] Cisco "Cisco Annual Internet Report (2018–2023)," 2020, [Online], https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

[5] I. Cvitić, D. Peraković, M. Periša, dan M. Botica, "Novel Approach for Detection of IoT Generated DDoS Traffic," *Wirel. Netw.*, Vol. 27, No. 3, hal. 1573–1586, Jun. 2021.

[6] A.W. Muhammad, C.F.M. Foozy, dan A. Azhari, "Machine Learning-Based Distributed Denial of Service Attack Detection on Intrusion Detection System Regarding to Feature Selection," *Int. J. Artif. Intell. Res.*, Vol. 4, No. 1, hal. 1–8, Jun. 2020.

[7] S.M. Kasongo dan Y. Sun, "Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset," *J. Big Data*, Vol. 7, hal. 1–20, Nov. 2020.

[8] M. Merouane, "An Approach for Detecting and Preventing DDoS Attacks in Campus," *Autom. Control Comput. Sci.*, Vol. 51, No. 1, hal. 13–23, Mar. 2017.

[9] Z. Ahmad, dkk., "Anomaly Detection Using Deep Neural Network for IoT Architecture," *Appl. Sci.*, Vol. 11, No. 15, Jul. 2021.

[10] O.F. Rashid, "DNA Encoding for Misuse Intrusion Detection System Based on UNSW-NB15 Data Set," *Iraqi J. Sci.*, Vol. 61, No. 12, hal. 3408–3416, Des. 2020.

[11] N. Moustafa dan J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set)," *2015 Mil. Commun., Inf. Syst. Conf. (MilCIS)*, 2015, hal. 1–6.

[12] M.S. Al-Daweri, K.A.Z. Ariffin, S. Abdullah, dan M.F.E.Md. Senan, "An Analysis of the KDD99 and UNSW-NB15 Datasets for the Intrusion Detection System," *Symmetry*, Vol. 12, No. 10, hal. 1-32, Okt. 2020.

[13] J.H. Corrêa, P.M. Ciarelli, M.R.N. Ribeiro, dan R.S. Villaça, "ML-Based DDoS Detection and Identification Using Native Cloud Telemetry Macroscopic Monitoring," *J. Netw. Syst. Manag.*, Vol. 29, No. 2, hal. 1–28, Jan. 2021.

[14] M. Tayyab, B. Belaton, dan M. Anbar, "ICMPV6-Based DOS and DDoS Attacks Detection Using Machine Learning Techniques, Open Challenges, and Blockchain Applicability: A Review," *IEEE Access*, Vol. 8, hal. 170529–170547, Sep. 2020.

[15] G.A. Jaafar, S.M. Abdullah, dan S. Ismail, "Review of Recent Detection Methods for HTTP DDoS Attack," *J. Comput. Netw. Commun.*, Vol. 2019, hal. 1–10, Jan. 2019.

[16] S. Rajagopal, K.S. Hareesha, dan P.P. Kundapur, "Feature Relevance Analysis and Feature Reduction of UNSW NB-15 Using Neural Networks on MAMLS," dalam *Advanced Computing and Intelligent Engineering. Advances in Intelligent Systems and Computing*, Vol. 1082, B. Pati, C. Panigrahi, R. Buyya, dan K.C. Li, Eds., Singapura, Singapura: Springer, 2018, hal. 321–332.

[17] A.M. Aleesa, M. Younis, A.A. Mohammed, dan N.M. Sahar, "Deep-Intrusion Detection System with Enhanced UNSW-NB15 Dataset Based on Deep Learning Techniques," *J. Eng. Sci. Technol.*, Vol. 16, No. 1, hal. 711–727, 2021.

[18] N. Moustafa dan J. Slay, "The Evaluation of Network Anomaly Detection Systems: Statistical Analysis of the UNSW-NB15 Data Set and the Comparison with the KDD99 Data Set," *Inf. Secur. J., A Global Perspec.*, Vol. 25, No. 1–3, hal. 18–31, Apr. 2016.

[19] A. Thakkar dan R. Lohiya, "A Survey on Intrusion Detection System: Feature Selection, Model, Performance Measures, Application Perspective, Challenges, and Future Research Directions," *Artif. Intell. Rev.*, Vol. 55, hal. 453-563, Jan. 2021.

[20] N. Kunhare dan R. Tiwari, "Study of the Attributes Using Four Class Labels on KDD99 and NSL-KDD Datasets with Machine Learning Techniques," *2018 8th Int. Conf. Commun. Syst., Netw. Technol. (CSNT)*, 2018, hal. 127–131.

[21] B. Bouyeddou, B. Kadri, F. Harrou, dan Y. Sun, "Nonparametric Kullback-Leibler Distance-Based Method for Networks Intrusion Detection," *2020 Int. Conf. Data Anal. Bus., Ind., Way Towards Sustain. Econ. (ICDABI)*, 2020, hal. 1–5.

[22] S. Khan, A. Gani, A.W.A. Wahab, dan P.K. Singh, "Feature Selection of Denial-of-Service Attacks Using Entropy and Granular Computing," *Arab. J. Sci. Eng.*, Vol. 43, No. 2, hal. 499–508, Feb. 2018.

[23] G.S. Kushwah dan V. Ranga, "Voting Extreme Learning Machine Based Distributed Denial of Service Attack Detection in Cloud Computing," *J. Inf. Secur. Appl.*, Vol. 53, hal. 1-12, Agu. 2020.

[24] J. Schmidt-Hieber, "The Kolmogorov–Arnold Representation Theorem Revisited," *Neural Netw.*, Vol. 137, hal. 119–126, Mei 2021.

[25] G. Kocher dan G. Kumar, "Analysis of Machine Learning Algorithms with Feature Selection for Intrusion Detection using UNSW-NB15 Dataset," *Int. J. Netw. Secur., Its Appl.*, Vol. 13, No. 1, hal. 21–31, Jan. 2021.

[26] M. Madhiarasan dan S.N. Deepa, "Comparative Analysis on Hidden Neurons Estimation in Multi Layer Perceptron Neural Networks for Wind Speed Forecasting," *Artif. Intell. Rev.*, Vol. 48, No. 4, hal. 449–471, Des. 2017.

[27] A. Sagheer, M. Zidan, dan M.M. Abdelsamea, "A Novel Autonomous Perceptron Model for Pattern Classification Applications," *Entropy*, Vol. 21, No. 8, hal. 1-24, Agu. 2019.