Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) DOI: 10.25126/jtiik.202183358
Vol. 8, No. 1, Februari 2021, hlm. 103-110 p-ISSN: 2355-7699
Akreditasi KEMENRISTEKDIKTI, No. 36/E/KPT/2019 e-ISSN: 2528-6579

METODE DETEKSI INTRUSI MENGGUNAKAN ALGORITME *EXTREME LEARNING MACHINE* DENGAN *CORRELATION-BASED FEATURE SELECTION*

Sulandri[*1], Achmad Basuki[2], Fitra Abdurrachman Bachtiar[3]

[1,2,3] Universitas Brawijaya Malang
Email: [1] sulandri_andri@student.ub.ac.id, [2] abazh@ub.ac.id, [3] fitra.bachtiar@ub.ac.id
[*]Penulis Korespondensi

(Naskah masuk: 11 Maret 2020, diterima untuk diterbitkan: 01 Februari 2021)

Abstrak

Deteksi intrusi pada jaringan komputer merupakan kegiatan yang sangat penting dilakukan untuk menjaga keamanan data dan informasi. Deteksi intrusi merupakan proses monitor *traffic* pada sebuah jaringan untuk mendeteksi adanya pola data yang dianggap mencurigakan, yang memungkinkan terjadinya serangan jaringan. Penelitian ini melakukan analisis pada *traffic* jaringan untuk mengetahui apakah paket tersebut mengandung intrusi atau merupakan paket normal. Data *traffic* yang digunakan untuk deteksi intrusi pada penelitian ini diambil dari *dataset* KDD Cup. Metode yang digunakan untuk melakukan deteksi intrusi dengan cara klasifikasi yaitu dengan menggunakan metode *Extreme Learning Machine* (ELM). Namun, dengan menggunakan metode ELM saja tidak mampu untuk menghasilkan akurasi yang baik maka, pada metode ELM perlu ditambahkan metode seleksi fitur *Correlation-Based Feature Selection* (CFS) untuk meningkatkan hasil akurasi dan waktu komputasi. Hasil penelitian yang dilakukan dengan menggunakan metode ELM menunjukkan tingkat akurasi mencapai 81,97% dengan waktu komputasi 3,39 detik. Setelah ditambahkan metode seleksi fitur CFS pada ELM tingkat akurasi meningkat secara signifikan menjadi 98,00% dengan waktu komputasi 2,32 detik.

Kata kunci: *keamanan jaringan, seleksi fitur, ELM, CFS*

INTRUSION DETECTION METHOD USING EXTREME LEARNING MACHINE ALGORITHM WITH CORRELATION-BASED FEATURE SELECTION

Abstract

Intrusion detection of computer networks is a very important activity carried out to maintain data and information security. Intrusion detection is the process of monitoring traffic on a network to detect any data patterns that are considered suspicious, which allows network attacks. This research analyzes the network traffic to find out whether the packet contains intrusion or is a normal packet. Traffic data used for intrusion detection in this study were taken from the KDD Cup dataset. The method used to do intrusion detection by classification is using the Extreme Learning Machine (ELM) method. However, using the ELM method alone is not able to produce good accuracy, so the ELM method needs to be added to the Correlation-Based Feature Selection (CFS) feature selection method to improve the accuracy and computational time. The results of the research conducted using the ELM method showed an accuracy rate of 81.97% with a computation time of 3.39 seconds. After adding the CFS feature selection method to ELM the accuracy level increased significantly to 98.00% with a computing time of 2.32 seconds.

Keywords: *network security, feature selection, ELM, CFS*

1. PENDAHULUAN

Sistem jaringan komputer harus dilindungi dari segala macam serangan baik dari dalam sistem maupun dari luar sistem, serta harus mampu mendeteksi usaha-usaha penyusupan oleh pihak yang tidak berhak. Keamanan jaringan komputer merupakan bagian terpenting dari sebuah sistem untuk menjaga informasi data serta menjamin ketersediaan layanan bagi pengguna. Oleh sebab itu sudah selayaknya pengembangan di bidang keamanan jaringan lebih diperhatikan, untuk mencegah kerusakan dari ancaman serangan yang saat ini semakin meningkat dan beragam.

Berdasarkan data yang diterbitkan oleh *Symantec* pada *Internet Security Threat Report* (ISTR) tahun 2017 (Security and Report, 2017) serangan terhadap keamanan jaringan semakin meningkat dari tahun ke tahun. Oleh sebab itu, diperlukan metode keamanan lalu lintas data informasi yang masuk kedalam sistem jaringan. Metode keamanan jaringan yang digunakan untuk medeteksi intrusi dalam mengatasi permasalahan tersebut, yaitu

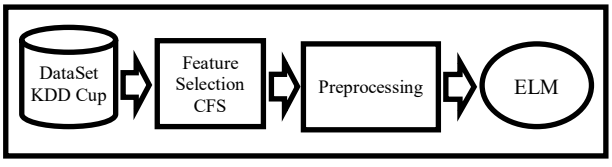
dengan menggunakan metode ELM dengan cara melakukan klasifikasi paket data. Metode ELM merupakan metode *supervised learning* yang dapat digunakan untuk mengklasifikasi dataset KDD Cup dengan tujuan layanan keamanan jaringan yang dapat melakukan pengawasan semua aktivitas di dalam sistem jaringan, serta mampu mengontrol dan mengevaluasi jenis paket data yang dianggap mencurigakan pada lalu lintas jaringan. Menurut (Neethu, 2003) dalam penelitiannya menilai bahwa perlu adanya terobosan baru untuk memperbaiki keterbatasan teknik pada keamanan jaringan, salah satu di antaranya menggunakan *Data Mining.*

Berbagai macam metode dan algoritme yang dapat digunakan dalam data mining untuk memecahkan permasalahan pada *cyber security* di antaranya yaitu *Extreme Learning Machine* (ELM). Menurut (Huang, Zhu and Siew, 2006) dalam penelitiannya, metode ELM merupakan jaringan *artificial neural* network dengan *single hidden layer* atau dapat disebut *single hidden layer feedforward neural networks* (SLFNs). Metode *supervised* ELM dapat digunakan untuk menutupi kelemahan dari jaringan *neural* network *feedforward* terutama dalam hal akurasi dan kecepatan pembelajaran.

Berdasarkan hasil evaluasi dari penelitian yang dilakukan (Benoît *et al.*, 2013) *Extreme Learning Machine* di samping mempunyai kelebihan juga terdapat beberapa kelemahan dalam klasifikasi. Hal ini diperkuat oleh (Abbas, Albadr and Tiun, 2017) pada penelitian yang dilakukan bahwa pada ELM jumlah *hidden neuron* ditentukan dengan cara melakukan banyak percobaan untuk mendapatkan informasi baru, sehingga tidak dapat diketahui besaran jumlah *hidden neuron* yang tepat untuk menghasilkan akurasi yang baik, hal ini mengakibatkan beberapa sampel kemungkinan akan terdapat rendahnya hasil klasifikasi pada kondisi tertentu. Oleh karena itu dalam penelitian yang dilakukan untuk membantu meningkatkan hasil klasifikasi dan efisiensi waktu komputasi yang selama ini menjadi permasalahan pada ELM. Dengan menambahkan algoritme seleksi fitur *Correlation-Based Feature Selection* (CFS) dengan tujuan untuk meningkatkan akurasi dan *leaning speed.*

2. METODE PENELITIAN

Metode pada penelitian ini yang terkait dengan serangan jaringan dengan menggunakan *Correlation-based Filter Selection* (CFS) dan *Extreme Learning Machine*. Tertulis informasi pada Gambar 1 mengenai langkah-langkah yang digunakan di dalam penelitian ini.

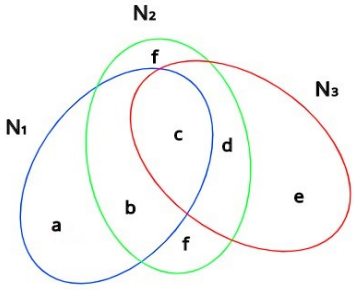


Gambar 1. Alur Penelitian

Alur proses dan penjelasan penelitian dapat dilihat pada gambar diatas. Tahap pertama merupakan tahapan pengolahan dataset KDD Cup agar dataset tersebut dapat diproses pada tahapan berikutnya. Tahapan kedua adalah proses seleksi fitur menggunakan metode CFS. Tahapan ketiga adalah *preprocessing* yang di dalamnya terdapat proses transformasi data, normalisasi dan pemetaan kelas serangan. Tahapan keempat adalah proses klasifikasi menggunakan algoritme ELM.

2.1. Seleksi Fitur Correlation-based Feature Selection (CFS)

Seleksi fitur berfungsi untuk menentukan suatu kelas pada nilai target dengan cara mengurangi jumlah fitur yang tidak relevan serta mengurangi dimensi data untuk meningkatkan performa sistem, efisiensi dan meningkatkan akurasi. Prinsip algoritme *Correlation-based Feature Selection* (CFS) seleksi fitur yang baik merupakan fitur yang relevan terhadap kelas akan tetapi fitur tersebut tidak *redudant* terhadap fitur-fitur yang lain, dengan menggunakan pendekatan *symmetrical uncertainty* (SU) untuk mengukur korelasi antara variable (Rodriguez dkk, 2019).



Gambar 2. Ilustrasi Seleksi Fitur dalam Diagram Veen (Sun, *dkk.*, 2017).

Fitur yang saling berkorelasi satu sama lain jika memiliki ukuran informasi timbal balik yang tinggi maka, fitur berisi banyak informasi tentang informasi yang lain dan dapat mewakili satu fitur yang lain, fitur yang sudah terwakili dapat tidak digunakan. Tahapan seleksi fitur yang diusulkan pada penelitian ini menggunakan *Corellation Feature Selection* (CFS). penelitian ini pada tahapan proses seleksi fitur akan dilakukan dengan menggunakan *tools* Weka. Weka adalah sebuah *tools* atau perangkat lunak yang dapat menerapkan berbagai algoritme machine learning untuk melakukan beberapa proses pada data mining misalnya seleksi fitur (Qu *dkk.*, 2018).

2.2. *Preprocessing*

Dataset yang digunakan untuk evaluasi serangan jaringan pada penelitian ini menggunakan KDD Cup, yang mana *dataset* tersebut dapat dibilang data yang masih mentah sehingga perlu dilakukan *preprocessing* data sebelum dilakukan proses klasifikasi dengan *machine learning*. Pada penelitian ini dilakukan *preprocessing* data pada tiga bagian yaitu transformasi

data, normalisasi dan penentuan tipe kelas serangan. Untuk memperoleh hasil klasifikasi dengan akurasi yang baik diperlukan data yang baik pula, sehingga mempermudah dalam pemilihan suatu metode yang akan digunakan. Tujuan dari *preprocessing* data adalah untuk mempersiapkan data agar data tersebut dapat meningkatkan efisiensi dan meningkatkan kualitas data. Metode yang digunakan dalam evaluasi pada penelitian ini untuk serangan jaringan menggunakan metode ELM dengan seleksi fitur CFS.

A. Transformasi Data

Dataset KDD Cup sebelum dilakukan proses klasifikasi menggunakan algoritme ELM-CFS terlebih dahulu dilakukan proses transformasi data. Maksud dan tujuan dari transformasi data adalah merubah bentuk data sehingga data siap untuk dianalisis. Permasalahan pada penelitian ini terletak pada *dataset* KDD Cup yang mana *dataset* tersebut terdapat data yang berupa *string,* maka perlu dilakukan perubahan data atau transformasi data untuk menjadi angka yaitu dengan cara memberikan penomoran terhadap data yang berupa *string*. Ada tiga parameter *dataset* KDD Cup yang harus dilakukan transformasi data seperti *service*, *flag* dan *protocol*_type.

B. Normalisasi

Tahapan normalisasi data merupakan tahapan yang sangat diperlukan, ketika data ada yang bernilai terlalu besar maupun data yang bernilai terlalu kecil, data tersebut akan kesulitan dalam memahami informasi data dalam proses klasifikasi. Tujuan normalisasi data yaitu untuk melakukan penskalaan fitur-fitur yang terdapat pada *dataset* KDD Cup, sehingga mempunyai batas atau jarak yang sama pada rentang tertentu serta untuk mengurangi *noise* data dan juga mengurangi dominasi nilai yang bernilai besar terhadap *variable* yang bernilai kecil (Singh D. and Singh B 2019). Perhitungan normalisasi dengan menggunakan rumus:

$$T' = \frac{T - min_a}{max_a - min_a} \quad (1)$$

T' merupakan data setelah dilakukan normalisasi, T data sebelum dilakukan normalisasi, min_a nilai minimal pada fitur i, max_a nilai maksimal pada fitur i.

C. Tipe Kelas Serangan Jaringan

Pengelompokan kelas berdasarkan kategori golongan serangan jaringan menjadi beberapa jenis serangan yang berfungsi untuk pengelompokan jenis serangan jaringan. Dataset KDD Cup dikelompokkan menjadi 5 kelas, terdapat 4 kelas yang merupakan kelas tipe serangan Misalnya serangan DoS terdapat beberapa jenis serangan (yaitu *land, back, smurf* dll). berikut klasifikasi tipe serangan pada jaringan menurut (Nskh, M, & Naik, 2016).

Tabel 1. Tipe Kelas Serangan Jaringan

Kelas	Tipe Serangan
DoS	neptune, pod, back, smurf, teardrop, land
U2R	perl, loadmodule, rootkit, buffer_overflow
R2L	guest_passwd, warezmaster, ftp_write, imap, multihop, phf, werezclient, spy.
PROBE	nmap, satan, ipsweep, portsweep
Normal	normal

2.3. *Extreme Learning Machine* (ELM)

Metode ELM diciptakan pertama kali oleh Huang. Metode ELM adalah metode Jaringan Saraf Tiruan (JST) dengan *feedforward* menggunakan *single hidden layer* atau yang sering disebut juga dengan Single Hidden Layer Feedforward Neural Networks (SLFNs). Metode pembelajaran ELM dibuat untuk mengatasi beberapa kekurangan dari jaringan saraf tiruan feedforward, terutama dalam proses waktu pembelajaran (Huang, Zhu & Siew, 2006). Susunan dari struktur metode ELM untuk proses klasifikasi, yaitu terdapat *input layer, hidden layer*, dan *output layer*. Penghitungan *matriks invers* menggunakan *Moore-Penrose* untuk menghitung keluaran atau *output hidden layer*. Dengan menggunakan *activation function sigmoid* karena *activation function* tersebut telah banyak dilakukan pengujian dengan menghasilkan akurasi terbaik pada banyak data (Cao *dkk.*, 2018).

ELM memiliki beberapa parameter seperti bobot *input,* bias dan *hidden neuron* dipilih dengan menggunakan cara secara acak, sehingga metode ELM memiliki waktu *learning* yang cepat dan mampu menghasilkan akurasi baik walaupun dengan menggunakan jumlah data yang besar.

Langkah-langkah *training* pada algoritme ELM dapat dilakukan dengan cara berikut ini (Cholissodin *et al.*, 2017):

1. Membuat nilai *random* bobot *W* (*weight*) dan bias dengan *range* tertentu. Dimana nilai acak matriks W_{jk} merupakan bobot *input* sedangkan nilai bias *b* dengan nilai matrik bias adalah [1*xj*] yang mana *k* adalah banyak node *input layer* dan *j* adalah banyak *hidden neuron*. *W* dan *b* merupakan parameter pembelajaran dari metode ELM.
2. Menghitung matriks *H* untuk nilai output diperoleh dari hidden layer dengan menggunakan fungsi aktivasi sigmoid.
   $$H = \frac{1}{(1+expexp\left(-(X_{training}.W^T + ones(N_{train},1) * bias)\right))} \quad (2)$$
   Dimana X adalah data untuk testing matriks dan W^T adalah matriks *transpose* bobot yang diperoleh dari *training*.
3. Menghitung matriks dengan menggunakan Moore-Penrose Generalized Inverse
   $$H^+ = (H^T.H)^{-1}.H^T \quad (3)$$
   H^+ merupakan *MoorePenrose Generalized invers matrik*s yang didapat dari matriks *H*. Sedangkan

matriks *H* adalah matriks yang terdiri dari keluaran masing *hidden* layer.

4. Melakukan perhitungan matriks bobot *output training* dari masukan *hidden* layer dengan menggunakan rumus sbb.

$$\hat{\beta} = H^T \; Y \quad (4)$$

$\hat{\beta}$ adalah hasil dari matriks keluaran bobot, *H+* adalah *Moore-Penrose* matriks.

5. Menghitung hasil prediksi

$$\hat{Y} = H \,.\hat{\beta} \quad (5)$$

Y merupakan target matrik prediksi, sedangkan *H* merupakan matriks keluaran dari *hidden* layer dan $\hat{\beta}$ adalah matriks keluaran bobot dari proses pembelajaran.

Langkah-langkah *testing* pada algoritme ELM dapat dilakukan dengan cara sebagai berikut:

1. Nilai bobot masukkan W_{jk}, $\hat{\beta}$ sesuai dengan pembelajaran dan nilai *b adalah* bias.
2. Mencari nilai *matrik H*

$$H_{test} = \frac{1}{(1+expexp\,(-(X_{training}.W^T+ones(N_{train},1)*bias)))} \quad (6)$$

3. Menghitung nilai dari hasil prediksi

$$\hat{Y} = H_{test} \,.\hat{\beta} \quad (7)$$

4. Menghitung nilai evaluasi. Nilai akurasi dan lama proess

2.4 *DATASET* KDD CUP

Dataset yang digunakan pada klasifikasi serangan jaringan untuk hasil evaluasi pada penelitian ini menggunakan *dataset* KDD Cup. *Dataset* ini merupakan data hasil pemantauan *traffic* yang dilakukan pada tahun 1999 dengan *tcpdump* yang digunakan untuk pendeteksian intrusi pada kegiatan "International *Knowledge Discovery and Data Mining Tools Competition*". jenis data serangan yang dikelompokkan ke dalam empat tipe intrusi dan satu data normal. Berikut persentase yang terdapat pada *dataset* KDD Cup.

Tabel 2. Persentase *Dataset* KDD Cup (Divekar, 2018)

Kelas	Data	Persentase
DoS	391458	79.24%
R2L	1126	0.23%
U2R	52	0.01%
Probe	4107	0.83%
Normal	97277	19.69%
Total	494020	100%

2.5. Metode Evaluasi

A. *Accuracy*

Akurasi adalah ukuran yang dapat digunakan untuk menilai ketepatan dari hasil uji coba yang telah dilakukan terhadap data sebenarnya yang diterapkan. Salah satu cara untuk mendapatkan nilai akurasi dengan membandingkan dari hasil deteksi yang telah dilakukan oleh sistem dengan data test sebenarnya yang digunakan. Pada penelitian ini akurasi diperoleh cara dengan membandingkan data yang telah berhasil dengan tepat diklasifikasikan ke dalam 5 kelas DoS, U2R, R2l, Probe dan Normal. Nilai akurasi dapat diperoleh dengan menggunakan rumus sebagai berikut.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

B. *Running time*

Running time adalah jumlah waktu yang digunakan pada proses pembelajaran dan pengujian yang dilakukan dengan tujuan akan didapatkan hasil akurasi. Terdapat beberapa penilaian yang sangat berpengaruh pada proses running time, yaitu adalah banyaknya jumlah data yang digunakan dan operasi dasar atau algoritme yang digunakan dalam proses klasifikasi. Satuan yang digunakan sebagai pengukuran dan penilaian dari *running time* dinyatakan dengan *seconds* (s).

4. ANCAMAN TERHADAP JARINGAN KOMPUTER

Beberapa kategori ancaman yang dapat membahayakan jaringan komputer adalah sebagai berikut:

A. *Denial-of-Service* (DoS)

DoS merupakan aktivitas yang bertujuan untuk menyerang komputer atau server dengan maksud membuat komputer korban tidak dapat memberikan layanannya terhadap para pengguna lain, yaitu dengan cara menghabiskan bandwidth pada sebuat *traffic*, sehingga kapasitas pemrosesan *router* atau *resource* tidak dapat menjalankan tugasnya dengan benar (Kumar dkk., 2014).

B. *Remote to Local* (R2L)

R2L merupan aktivitas serangan komputer dengan untuk mencari celah keamanan dari suatu computer lain atau jaringan yang menjadi target dengan maksud untuk memperoleh akses terhadap komputer target yang dituju dengan melalui remote jaringan komputer (Hassan, 2017).

C. *User to Root* (U2R)

U2R adalah aktivitas dari pengguna *anonym* atau yang sering disebut pengguna normal dengan tujuan untuk memperoleh hak akses dari komputer korban sebagai *administrator* (*root* atau *super user*) dengan cara melakukan *exploits* untuk pemanfaatan memperoleh keuntungan terhadap komputer atau jaringan (Hassan, 2017).

D. *Probing*

Probing adalah tindakan yang membahayakan dalam jaringan komputer dengan maksud dan tujuan untuk mencari kelemahan-kelemahan atau celah keamanan suatu sistem komputer maupun jaringan dengan cara melakukan pengawasan secara intensif

terhadap target. Teknik atau metode untuk melakukan probing dapat disebut juga dengan probe (Ahmad and Iskandar, 2009).

4. PERANCANGAN SISTEM

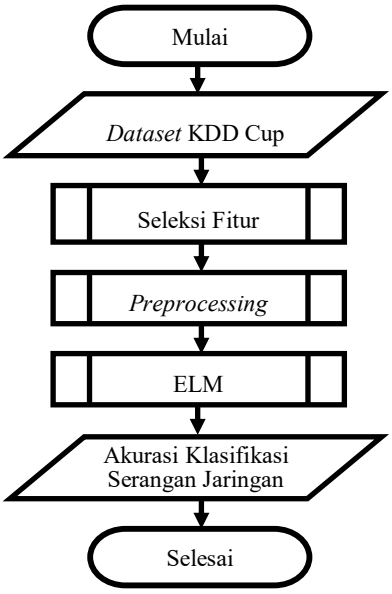
Perancangan merupakan gambaran sebuah sistem yang dibentuk untuk mendefinisikan kebutuhan-kebutuhan yang diperlukan dalam penelitian.

4.1. Pengumpulan Data

Data yang digunakan berdasarkan pengolahan data DARPA yaitu KDD cup. Data KDD merupakan data pengujian yang banyak digunakan untuk menganalisis algoritme *Data mining* yang menurut penelitian sebelumnya merupakan algoritme terbaik pada kasus klasifikasi untuk deteksi anomali. Data pada *dataset* KDD Cup terdiri dari data serangan dengan fitur-fitur yang bisa digunakan untuk mendeteksi serangan. *Dataset* KDD Cup terdiri dari 41 fitur, serta terdapat 1 label *class* fitur tambahan.

4.2. Alur Perancangan Sistem

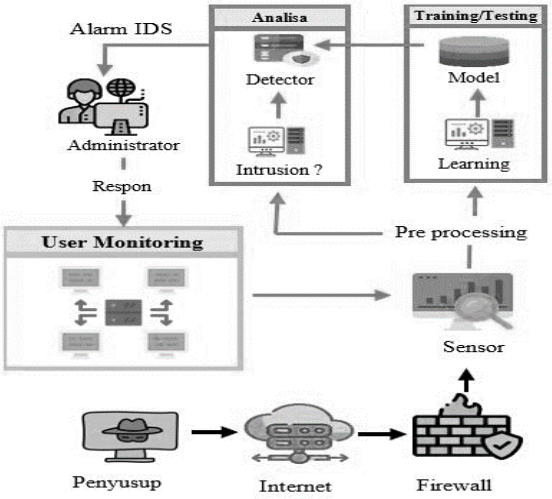
Proses awal yang dilakukan pada penelitian ini dimulai dari pengolahan data set dengan cara seleksi fitur dengan tujuan untuk membuang data yang tidak sesuai dan mengurangi dimensi data. Selanjutnya akan dilakukan transformasi data atau perubahan data dari bentuk string menjadi angka. Setelah dilakukan seleksi fitur dan transformasi data tahapan selanjutnya yaitu adalah melakukan penskalaan dengan teknik normalisasi untuk mendapatkan *value* dari rentang nilai antara -1 hingga 1. Tahapan selanjutnya dilakukan klasifikasi menggunakan ELM dengan cara melakukan proses pembelajaran dan pengujian untuk mendapatkan akurasi dari klasifikasi serangan jaringan komputer.



Gambar 4. Alur Perancangan Sistem

4.3. Skema Proses Deteksi Serangan

Mekanisme deteksi aktivitas yang mencurigakan pada sebuah jaringan dengan memanfaatkan *machine learning* pada penelitian ini dapat dilihat pada gambar 5. Penelitian ini bertujuan untuk membuat sebuah *prototype* untuk deteksi intrusi yang memiliki kemampuan mencari data yang dianggap sebagai serangan jaringan dari banyaknya data yang telah dilakukan pembelajaran atau *training*.



Gambar 5. Skema Proses Deteksi Serangan Jaringan

5. HASIL DAN PEMBAHASAN

Implementasi yang dilakukan dengan spesifikasi PC Intel Core i7, RAM 8 GB, dan 1 TB HDD, diperoleh hasil pengujian terhadap nilai *hidden neuron* dan waktu komputasi. Perbandingan data yang digunakan dalam proses pembelajaran dan pengujian sebesar 80% untuk training dan 20% untuk pengujian. Pengujian pertama dilakukan menggunakan algoritme ELM menggunakan seleksi fitur, pada setiap pengujian dilakukan sebanyak lima kali pengujian dari setiap hidden neuron dengan masing-masing pengujian menggunakan fungsi aktivasi sigmoid. Pengujian kedua dilakukan dengan menambahkan seleksi fitur pada ELM. Nilai akurasi dan waktu komputasi pada penelitian ini diambil rata-rata dari setiap lima kali pengujian.

5.1. Pengujian Hidden Neuron pada Nilai Akurasi

Nilai pada *hidden neuron* dengan menggunakan nilai bebas yang digunakan pada *Extreme Learning Machine* untuk memperoleh nilai akurasi. Pengujian terhadap nilai *hidden neuron* dilakukan dengan nilai 5, 10, 15, 20, 30, 35, 40, 45, 50, 60, 70 dan 100. Masing-masing hidden neuron yang telah disebutkan diatas dilakukan pengujian sebanyak 5 kali dan didapatkan nilai akurasi berdasarkan rata-rata tiap hidden neuron. Dalam pengujian *hidden neuron* menggunakan fungsi aktivasi sigmoid.

A. Pengujian CFS-ELM

Proses seleksi fitur pada penelitian yang dilakukan dengan menggunakan bantuan sebuah *tools* yaitu Weka.

Weka menyediakan beberapa algoritma seleksi fitur salahsatu diantaranya adalah seleksi fitur CFS. Hasil dari proses seleksi fitur menggunakan *tools* Weka menunjukan bahwa dari 41 fitur yang terdapat pada *Dataset* KDD Cup terseleksi menjadi 11 fitur, dengan nilai korelasi antara fitur yang sangat rendah tetap memiliki nilai dengan korelasi yang tinggi antara fitur dan kelas. Fitur yang terpilih dari proses *feature selection* ini dengan menggunakan CFS adalah fitur *{dst_host_ same_ src_port_rate, protocol_type, diff_srv_ rate, service, count, flag, root_shell, src_bytes, wrong_fragment, dst_byte, land}*.

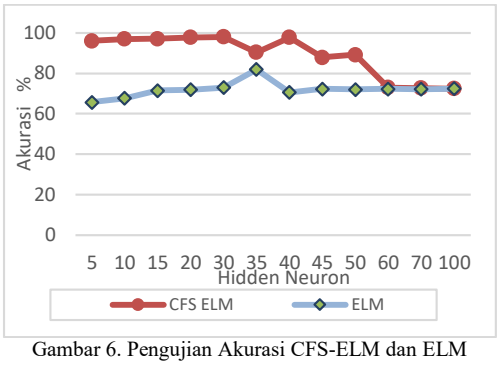
Tabel 3. Pengujian CFS-ELM

HN	Pengujian					AV	T
	1	2	3	4	5		
5	95,8	96	96,2	95,9	96,3	96	0,31
10	97	97,1	96,9	97,1	96,6	97	0,59
15	97,2	96,6	96,6	97,7	97,4	97,1	1,07
20	97,6	98	97,7	97,7	97,9	97,8	1,37
30	98	98	98,1	98	97,9	98	2,32
35	72,7	98	97,9	98	85,2	90,4	2,81
40	98,1	98	97,4	97,7	97,7	97,8	3,31
45	98	98,1	72,6	98,1	72,7	87,9	3,72
50	72,8	78,6	98	98,1	98,1	89,1	4,3
60	72,8	72,8	73,2	72,6	73,1	72,9	5,31
70	72,6	72,9	72,2	73,1	72,5	72,7	6,79
100	72,7	72,4	72,2	72,4	72,3	72,4	11,2

Terlihat pada Tabel 3. HN merupakan *Hidden Neuron* yang digunakan, AV merupakan rata-rata hasil dari lima kali pengujian dan T adalah rata-rata waktu pengujian. Diperoleh hasil akurasi deteksi pada 5 kali percobaan dari metode ELM menggunakan seleksi fitur CFS dengan nilai rata-rata akurasi tertinggi dengan nilai 98,00% dan jumlah *hidden neuron* yang digunakan sebanyak 30, sedangkan hasil akurasi terendah terdapat pada *hidden neuron* 100 dengan nilai akurasi 72,40. Dalam setiap proses pengujian menggunakan fungsi aktivasi sigmoid.

B. Pengujian ELM

Pengujian dilakukan dimulai dari nilai *hidden neuron* yang terkecil yaitu dengan 5 *hidden neuron* memperoleh rata-rata hasil akurasi 65,64%. Pengujian dilakukan hingga 100 *hidden neuron* yang terbanyak dengan memperoleh rata-rata akurasi 72,4%. Dari pengujian yang telah dilakukan nilai rata-rata akurasi tertinggi menggunakan 35 hidden neoron rata-rata akurasi 81,9%. Hasil evaluasi pengujian dapat dilihat pada Tabel 4.



Gambar 6. Pengujian Akurasi CFS-ELM dan ELM

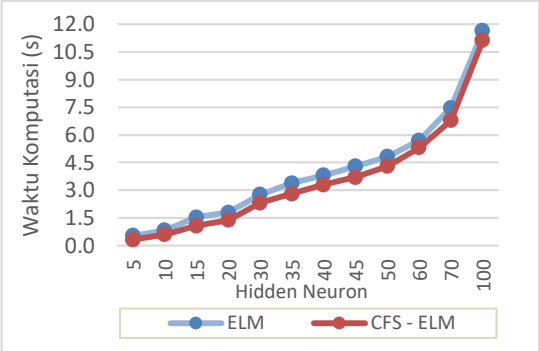
Tabel 4. Pengujian ELM

HN	Pengujian					AV	T
	1	2	3	4	5		
5	47	82,9	92,3	48,8	57,2	65,6	0,54
10	67,5	56,3	70,4	79	65,3	67,7	0,83
15	90,5	70,4	65,6	70,7	60,3	71,5	1,54
20	71	62	64,9	69,1	92,7	72	1,8
30	72,1	71,7	67,9	84,1	69,1	73	2,77
35	70,5	94,7	95,9	76,9	71,9	81,9	3,39
40	72,9	69,7	71	70,2	69,4	70,6	3,81
45	72	71,2	70,9	77	70,2	72,3	4,3
50	71,8	72,1	72,1	72,1	72	72	4,82
60	71,9	72,4	72,2	72,2	72,4	72,2	5,71
70	71,7	72,1	72,5	72,2	72,3	72,1	7,47
100	72,6	72,4	72,4	72,1	72,4	72,4	11,7

5.2. Pengujian *Hidden neuron* pada Nilai Waktu Komputasi

Evaluasi pengujian dilakukan bertujuan untuk melihat seberapa besar pengaruh jumlah *hidden neuron* yang telah diimplementasikan terhadap waktu proses CFS-ELM maupun ELM. Berdasarkan hasil evaluasi pengujian yang telah dilakukan, penggunaan dengan jumlah *hidden neuron* yang terkecil menghasilkan waktu komputasi tercepat. Dapat dilihat pada gambar 7, bahwa algoritme CFS-ELM dengan menggunakan 5 *hidden neuron* membutuhkan waktu komputasi 0,31 detik, sedangkan dengan jumlah *hidden neuron* 100 waktu komputasi 11,15 detik.

Pengujian pada algoritme ELM membutuhkan waktu lebih banyak dibandingkan dengan CFS- ELM, dengan jumlah 5 *hidden neuron* waktu komputasi sebesar 0,54 detik. Terdapat selisih waktu 0,23 detik apabila menggunakan CFS-ELM. Dalam keseluruhan pengujian apabila diambil rata-rata maka, terdapat selisih waktu 5,60 detik. Untuk lebih jelasnya dapat kita lihat pada gambar 7.



Gambar7. Perbandingan Waktu Komputasi CFS-ELM dengan ELM

6. KESIMPULAN

Hasil evaluasi menggunakan metode ELM dengan seleksi fitur CFS menunjukkan peningkatan hasil akurasi yang signifikan. Keseluruhan dari pengujian yang dilakukan menggunakan fungsi aktivasi sigmoid. Hasil evaluasi menggunakan metode seleksi fitur CFS pada ELM dengan 5 kali pengujian menunjukkan akurasi tertinggi dengan nilai rata-rata sebesar 98,00% dengan jumlah *hidden neuron* 30 dan rata-rata waktu proses komputasi sebesar 2,32 detik. Sedangkan pengujian menggunakan metode ELM tanpa seleksi fitur nilai akurasi sebesar 81,97% dengan jumlah *hidden neuron* 35 dan waktu komputasi sebesar 3,39 detik. Terbukti dari hasil evaluasi yang dilakukan dengan menggunakan seleksi fitur tingkat akurasi meningkat sebesar 16,36% dan waktu komputasi lebih singkat 1,07 detik.

Penelitian yang dilakukan menggunakan algoritme ELM dengan seleksi fitur CFS dapat diimplementasikan dalam intrusi deteksi serangan jaringan komputer dengan cara mengukur tingkat akurasi. Akurasi yang didapat dari evaluasi yang telah dilakukan menunjukkan bahwa algoritme ELM dengan seleksi fitur CFS lebih unggul dibandingkan dengan metode ELM tanpa seleksi fitur. Dari segi waktu proses komputasi lebih cepat menggunakan algoritme ELM dengan seleksi fitur CFS.

DAFTAR PUSTAKA

ABBAS, M., ALBADR, A. & TIUN, S. 2017 'Extreme Learning Machine: A Review', 12(14), pp. 4610–4623.

AHMAD, I. & ISKANDAR, B. S. 2009 'Application of Artificial Neural Network in Detection of Probing Attacks', *2009 IEEE Symposium on Industrial Electronics & Applications*. IEEE, 2(Isiea), pp. 557–562. doi: 10.1109 /ISIEA.2009. 5356382.

BENOÎT, F. *dkk.* 2013 'Feature Selection for Nonlinear Models with Extreme Learning Machines', *Neurocomputing*, 102, pp. 111–124. doi: 10.1016 /j.neucom.2011.12. 055.

CAO, J. *dkk.* 2018 'Extreme Learning Machine with Affine Transformation Inputs in an Activation Function', *IEEE Transactions on Neural Networks and Learning Systems*. IEEE, PP(November), pp. 1–15. doi: 10.1109/ TNNLS.2018.2877468.

CHOLISSODIN, I. *dkk.* (2017) 'Optimasi Kandungan Gizi Susu Kambing Peranakan Etawa (PE) Menggunakan Elm-Pso Di Upt Pembibitan Ternak Dan Hijauan', 4(1), pp. 31–36.

DIVEKAR, A. (2018) 'Benchmarking datasets for Anomaly-based Network Intrusion Detection : KDD CUP 99 alternatives', *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*. IEEE, pp. 1–8.

HASSAN, D. 2017 'Cost-Sensitive Access Control for Detecting Remote to Local (R2L) and User to Root (U2R) Attacks', 43(2), pp. 124–129.

HUANG, G. BIN, ZHU, Q. Y. & SIEW, C. K. 2006 'Extreme Learning Machine: Theory and Applications', *Neurocomputing*, 70(1–3), pp. 489–501. doi: 10.1016/j.neucom.2005. 12.126.

KUMAR, G. 2014 'Understanding Denial of Service (DoS) Attacks Using OSI Reference Model', (5), pp. 10–17.

KUMAR, S. *dkk.* 2014 'A Detail Analysis on Intrusion Detection Datasets', (May). doi: 10.1109/IAdCC .2014.6779523.

LIAO, H. J. *dkk.* 2013 'Intrusion Detection System: A Comprehensive Review', *Journal of Network and Computer Applications*. Elsevier, 36(1), pp. 16–24. doi: 10.1016 /j.jnca.2012.09.004.

NSKH, P., M, N. V. & NAIK, R. R. 2016 'Principle Component Analysis based Intrusion Detection System Using Support Vector Machine', *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. IEEE, pp. 1344–1350. doi: 10.1109 /RTEICT.2016.7808050.

RODRIGUEZ, D. (2019) 'Distributed Correlation-Based Feature Selection in Spark', pp. 1–25. *Information Sciences*. Elsevier https://doi.org /10.1016/j.ins. 2018.10.052

QU, D. *dkk.* 2018 'Journal of Network and Computer Applications A cache-aware social-based QoS routing scheme in Information Centric Networks', *Journal of Network and Computer Applications*. Elsevier Ltd, 121(January), pp. 20–32. doi: 10.1016/j.jnca. 2018.07.002.

SECURITY, I. & Report, T. n.d [online] 2017 'Internet Security Treat Report' Tersedia di <https://www.symantec.com/content/dam/symantec/docs/reports/istr-22-2017-en.pdf>, 22 (April).

SINGH, D. DAN SINGH, B. 2019 'Investigating the Impact of Data Normalization On Classification Performance', *Applied Soft Computing Journal*. Elsevier B.V., p. 105524. doi: 10.1016/j.asoc.

2019.105524.

SUN, Y. *dkk.* (2017) 'Correlation Feature Selection and MutualInformation Theory Based Quantitative Researchon Meteorological Impact Factors of ModuleTemperature for Solar Photovoltaic Systems '. doi: 10.3390/en10010007.