p-ISSN: 2301-5373
e-ISSN: 2654-5101

Klasifikasi Serangan *Distributed Denial of Service* (DDoS) Menggunakan *Random Forest* dengan CFS

I Made Wasanta Bhaskara[a1], I Putu Gede Hendra Suputra[a2], I Made Widiartha[a3], I Gusti Agung Gede Arya Kadyanan[a4], I Gusti Ngurah Anom Cahyadi Putra[a5], Ida Bagus Gede Dwidasmara[a6]

[a]Informatics Engineering, Faculty of Math and Science, University of Udayana
South Kuta, Badung, Bali, Indonesia
[1]wasanta.bhaskara32@gmail.com
[2]hendra.suputra@unud.ac.id
[3]madewidiartha@unud.ac.id
[4]gungde@unud.ac.id
[5]anom.cp@unud.ac.id
[6]dwidasmara@unud.ac.id

Abstract

Distributed Denial of Service (DDoS) attacks can have serious impacts on your organization and can cause enormous losses. This attack works by sending a computer or server an amount of requests that exceeds the capabilities of that computer. When classifying DDoS attacks in this study, feature selection is performed using correlation-based feature selection (CFS). The dataset used by the author in this study is CSE-CIC-IDS 2018. Feature selection on a dataset using CFS gets the results in the form of features related to the dataset. That is, a total of 31 features with a relationship score greater than 0.1. The average precision generated by the system using the random forest method and CFS function selection is 99.784%. Accuracy is the result of using the number of trees parameter with a value of 10. For a random forest model with no feature selection, the highest accuracy is 49.501%. This indicates that changing the random forest model parameters and selecting the CFS feature will affect high accuracy.

Keywords: *Distributed Denial of Service (DDoS), Correlation-based Feature Selection, Random Forest, Classification, CSE-CIC-IDS2018, Feature Selection*

1. Pendahuluan

Perkembangan teknologi yang semakin pesat meningkatkan ketergantungan masyarakat terhadap penggunaan sistem komputer. Sistem komputer digunakan dalam berbagai aspek kehidupan seperti aspek keuangan, kesehatan, industri dan aspek lainnya. Sistem komputer tersebut umumnya saling terhubung satu sama lain dengan menggunakan berbagai macam jaringan. Hal itu tentu memungkinkan terjadinya risiko keamanan dalam sistem jaringan komputer. Dalam beberapa tahun terakhir, seluruh dunia telah menyaksikan sejumlah besar insiden jaringan. Beberapa dari peristiwa jaringan ini diatasi menggunakan tindakan yang dapat mendeteksi serangan jaringan dalam rangka peningkatan keamanan jaringan komputer.

Tantangan deteksi serangan jaringan saat ini adalah semakin beragamnya lalu lintas dan jenis serangan baru. Salah satu jenis serangan pada jaringan internet yang paling sering terjadi yaitu *Distributed Denial of Service* (DDoS). Serangan *Distributed Denial of Service* (DDoS) memiliki efek yang serius terhadap suatu perusahaan dan dapat 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 [1]. Serangan DDoS mampu melumpuhkan server dengan membanjiri lalu lintas jaringan dan mengakibatkan *server down*. Ancaman dan serangan terhadap keamanan *server* terus meningkat, banyaknya kemudahan dan ketersedian informasi mengenai hacking yang dapat diakses dengan mudah di internet sehingga menjadikan pelaku mudah mendapatkan informasi untuk dijadikan sebagai target kejahatan [2]. *Attacker* (penyerang) dapat melakukan *crack* pada beberapa mesin yang nantinya komputer utama akan mengendalikan beberapa mesin menjadi *zombie-zombie* (botnet) yang secara terdistribusi menyerang server tujuan untuk meniadakan ketersedian informasi dari korban. Serangan *Distributed Denial of Service* (DDoS)

menyebabkan host atau jaringan tidak dapat menerima atau memproses *request* normal secara tepat waktu sehingga gagal memberikan layanan normal kepada pengguna.

Penelitian yang dilakukan oleh Sofa & Subiyanto (2020) [3] membahas *Smart Intrusion Detection System* berbasis *Compression Header Analyzer* dalam menganalisis varian baru model routing attacks yang terdapat pada Internet of Things dengan machine learning algorithms yang digunakan untuk klasifikasi *routing attacks* ada 6, antara lain *Random Forest*, J48, *Logistic*, MLP, *Naïve Bayes*, dan SMO. Berdasarkan hasil penelitian, Dari enam machine learning algorithm tersebut, kinerja terbaik dalam mendeteksi *routing attacks* ditunjukkan oleh *Random Forest* dengan tingkat akurasi sebesar 99,4721%. Selain itu, *Random Forest* juga mencapai nilai *True Positive* (TP) tertinggi sebesar 0,995 dan nilai *Mean Absolute Error* (MAE) terendah sebesar 0,0008. Sehingga berdasarkan penelitian tersebut, penulis akan menggunakan klasifikasi *Random Forest* dengan jenis serangan yang berbeda dari routing attacks untuk membuktikan apakah *Random Forest* juga akan memiliki tingkat akurasi yang tinggi pada jenis serangan *Distributed Denial of Service* (DDoS).

Seleksi fitur merupakan salah satu teknik penting untuk dilakukan dalam preprocessing data. Proses seleksi fitur bertujuan untuk menentukan jumlah fitur yang akan digunakan dalam menentukan kelas target serta mengurangi fitur yang tidak relevan. Pada penelitian yang dilakukan oleh Kurniabudi et al (2020) [4] membahas mengenai seleksi fitur dengan *Information Gain* untuk meningkatkan deteksi serangan *Distributed Denial of Service* (DDoS) menggunakan *Random Forest*. Pada penelitian tersebut menggunakan seleksi fitur *Information Gain* yang merupakan teknik *filtered-based* terhadap dataset CICIDS-2017. Berdasarkan data hasil eksperimen, dapat disimpulkan bahwa teknik seleksi fitur *Information Gain* mampu meningkatkan performa metoda klasifikasi khususnya *Random Forest* yang memiliki performa yang lebih baik dibandingkan *Naïve Bayes*. Selanjutnya penelitian yang dilakukan oleh Asmoro et al (2018) [5] membahas perbandingan kinerja seleksi fitur antara *Correlation-based Feature Selection* (CFS) dengan *Information Gain* pada prediksi kinerja akademik siswa dengan metode C4.5. Teknik seleksi fitur *Correlation-based Feature Selection* (CFS) merupakan algoritma seleksi fitur yang paling stabil dari semua pengujian skala perankingan tingkat densitas data, sedangkan algoritma seleksi fitur yang berbasis *entropy* mempunyai kecenderungan dalam memilih atribut yang sama dengan jumlah yang sama. Dari hasil penelitian, data yang diuji menggunakan algoritma C4.5 yang dikombinasikan dengan seleksi fitur CFS menghasilkan nilai akurasi yang lebih tinggi yaitu sebesar 76,92% dibandingkan dengan seleksi fitur *Information Gain* yang hanya memiliki nilai akurasi sebesar 76,19%. Maka dari itu penulis akan menggunakan seleksi fitur *Correlation-based Feature Selection* (CFS) untuk menguji apakah seleksi fitur *Correlation-based Feature Selection* (CFS) menggunakan metode *Random Forest* dapat meningkatkan akurasi atau performa *Random Forest* dibandingkan dengan *Information Gain* menggunakan metode *Random Forest*.

Random Forest dapat meningkatkan akurasi karena adanya pemilihan secara acak dalam membangkitkan simpul anak untuk setiap node (simpul di atasnya) dan diakumulasikan hasil klasifikasi dari beberapa pohon (tree). Keputusan akhir diambil dari hasil klasifikasi yang paling banyak muncul. Dalam pengklasifikasian serangan *Distributed Denial of Service* (DDoS) pada penelitian ini akan dilakukan seleksi fitur menggunakan *Correlation-based Feature Selection* (CFS). Dataset yang akan digunakan penulis dalam penelitian ini adalah CSE-CIC-IDS2018. Sebelum menguji dataset, terlebih dahulu akan dilakukan seleksi fitur sehingga mendapatkan fitur yang paling relevan dalam hasil data uji untuk mendapatkan hasil yang lebih akurat. Setelah mendapatkan fitur-fitur tersebut maka data uji akan diklasifikasikan menggunakan metode *Random Forest* sehingga dapat mengetahui pola serangan *Distributed Denial of Service* (DDoS) dan dapat mencegah serangan *Distributed Denial of Service* (DDoS) lebih awal.

2. Metode Penelitian

Pada penelitian ini menggunakan data riset CSE-CIC-IDS2018 yang didapat pada situs website www.unb.ca. Dalam penelitian ini menggunakan beberapa data dari dataset CSE-CIC-IDS2018 yang memiliki total 1.048.575 record data dan terdiri atas data normal dan data serangan DDoS yang dikelompokan kedalam jenis serangan DDoS HOIC dan DDoS LOIC UDP. Namun, dalam penelitian ini penulis hanya menggunakan 2 kelas yaitu normal dan serangan DDoS HOIC (*Distributed Denial of Service*). Jumlah paket data yang terdapat pada dataset yang penulis gunakan berjumlah 3 yaitu Normal, DDoS HOIC dan DDoS LOIC UDP dan dataset CSE-CIC-IDS2018 terdiri dari 80 fitur.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2.1 Data Mining

Data mining dapat diartikan sebagai proses mengekstrak atau menggali knowledge yang ada pada sekumpulan data. Informasi dan *knowledge* yang didapat tersebut dapat digunakan pada banyak bidang, seperti manajemen bisnis, pendidikan, kesehatan dan sebagainya. Data *mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* yang besar. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Proses menggali informasi dalam data mining melibatkan integrasi teknik dari berbagai disiplin ilmu, seperti teknologi *database* dan data *warehouse*, statistik, *machine learning*, komputasi dengan kinerja tinggi, *pattern recognition*, *neural network*, visualisasi data dan sebagainya.

Secara garis besar, data mining dapat dikelompokkan menjadi 2 kategori utama, yaitu:

a. *Descriptive mining*, yaitu proses untuk menemukan karakteristik penting dari data dalam satu basis data. Teknik data mining yang termasuk descriptive mining adalah *clustering*, *association*, dan *sequential mining*.
b. *Predictive mining*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi.

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. *Description*

   Deskripsi adalah teknik yang dilakukan peneliti dan analis secara sederhana dalam mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. *Classification*

   Klasifikasi adalah teknik yang paling umum diterapkan pada data mining. Klasifikasi merupakan proses menemukan sebuah model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefinisikan sebelumnya. Dalam klasifikasi pengujian data dilakukan dengan menggunakan perkiraan akurasi dari aturan klasifikasi. Jika akurasi bisa diterima, maka aturan dapat diterapkan untuk data baru. Salah satu contoh yang mudah dan populer adalah dengan *decision tree*.

3. *Clustering*

   Clustering bisa dikatakan sebagai identifikasi kelas objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Dengan teknik *clustering* kita bisa lebih lanjut mengidentifikasi dan menemukan secara keseluruhan pola distribusi dan korelasi antara atribut. Pendekatan klasifikasi secara efektif juga dapat digunakan untuk membedakan kelompok atau kelas objek.

4. *Prediction*

   Prediksi hampir sama dengan klasifikasi dan estimasi, perbedaan nya adalah dalam prediksi nilai dari hasil akan ada dimasa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan untuk keadaan yang bisa diprediksi. Salah satu contoh prediksi adalah prediksi harga beras dalam tiga bulan yang akan datang.

5. *Association rule*

   Tugas dari asosiasi adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Tugas asosiasi berusaha untuk mengungkap aturan untuk mengukur hubungan antara dua atau lebih atribut.

2.2 Correlation-based Feature Selection (CFS)

Seleksi fitur *correlation-based* merupakan proses seleksi fitur dengan melakukan perhitungan dan perbandingan tingkat korelasi antara atribut dengan kelas dan atribut dengan atribut lainya. Atribut

yang dipilih merupakan atribut memiliki korelasi tinggi dengan kelasnya dan memiliki tingkat korelasi rendah dengan atribut lainya [5]. Dengan rumus persamaan:

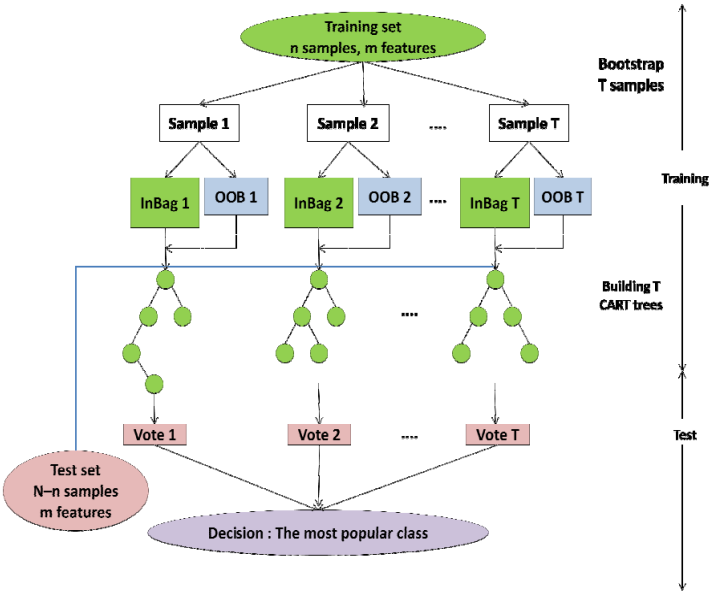
$$r_{sk} = \frac{k\bar{r}_{cf}}{\sqrt{k+k(k-1)r_{ff}}} \quad (1)$$

Dimana r_{sk} adalah hubungan antara fitur, k adalah jumlah fitur dan $\bar{r}_{cf}$ merupakan rata-rata hubungan fitur dan kelas, sedangkan r_{ff} adalah rata-rata korelasi antara bagian fitur [6].

2.3 *Random Forest*

Algoritma *Random Forest* merupakan *ensemble classifier* jenis bagging dimana model ini terdiri dari beberapa decision tree dan kelas *output* yang dihasilkan yaitu kelas dari *individual tree* [1]. *Random Forest* merupakan metode klasifikasi yang *supervised*. Sesuai dengan Namanya, metode ini menciptakan sebuah hutan (*forest*) dengan sejumlah pohon (*tree*). Secara umum, semakin banyak pohon (*tree*) pada sebuah hutan (*forest*), maka semakin kuat juga hutan tersebut. Pada kasus yang sama, semakin banyak *tree*, maka semakin besar pula akurasi yang didapatkan [7].

Metode *Random Forest* dilakukan dengan membangun pohon keputusan (*tree*) yang terdiri dari *root node*, *internal node*, dan *leaf node* dengan mengambil atribut dan data secara acak sesuai ketentuan yang diberlakukan. *Root node* merupakan simpul yang terletak paling atas, atau biasa disebut sebagai akar dari pohon keputusan. *Internal node* adalah simpul percabangan, dimana *node* ini mempunyai *output* minimal dua dan hanya ada satu *input*. Sedangkan *leaf node* atau *terminal node* merupakan simpul terakhir yang hanya memiliki satu *input* dan tidak mempunyai *output* [8].



Gambar 1. Struktur Random Forest (Sumber: Guo, 2011)

Pada saat proses *bootstrapping*, sebanyak sepertiga dari sampel akan digunakan untuk menentukan data *Out of Bag* (OOB). Perhitungan OOB *Error* dilakukan dengan tujuan untuk mencari *Out of Bag* pada data, dimana OOB *Error* merupakan data yang tidak termuat dalam *bootstrap* pada *random forest*. Setelah membangun pohon dengan n sampel *bootstrap*, maka dapat dilakukan pengujian menggunakan masing-masing data sampel yang ditinggalkan dengan menghitung kesalahan prediksi rata-rata dari sampel tersebut. Perhitungan skor OOB dapat dilakukan untuk setiap pohon dan diambil rata-ratanya dari semua skor tersebut untuk mendapatkan perkiraan seberapa akurat kinerja *random forest* yang telah dibuat. Hasil OOB *error* akan memberikan perkiraan seberapa akurat model *random forest* tanpa harus menguji secara formal dengan dataset yang baru. Dengan perhitungan OOB *error* dapat diketahui perkiraan tingkat kesalahan yang sama yang akan didapatkan pada waktu pengujian.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2.4 Confusion Matrix

Untuk melakukan evaluasi, dilakukan pengukuran akurasi metode *Random Forest* menggunakan *Confusion Matrix*. Perhitungan menggunakan *Confusion Matrix* didapatkan dari TP, TN, FP, dan FN [1].

- *False Positive* (FP), adalah catatan yang serangan tetapi diklasifikasikan sebagai catatan normal.
- *False Negative* (FN), adalah catatan yang normal tetapi diklasifikasikan sebagai catatan serangan.
- *True Positive* (TP), adalah jumlah catatan yang normal dan diklasifikasikan sebagai catatan normal.
- *True Negative* (TN), adalah jumlah catatan yang serangan dan diklasifikasikan sebagai catatan serangan.

Tabel 1. Confusion Matrix

Aktual	Kelas Prediksi		Jumlah
	Normal	Serangan	
Normal	TP	FN	P
Serangan	FP	TN	N
Jumlah	P'	N'	P + N

Adapun rumus untuk mengukur akurasi sebagai berikut:

$$Accuracy = \frac{TP+TN}{P+N} \quad (2)$$

2.5 Skenario Pengujian Sistem

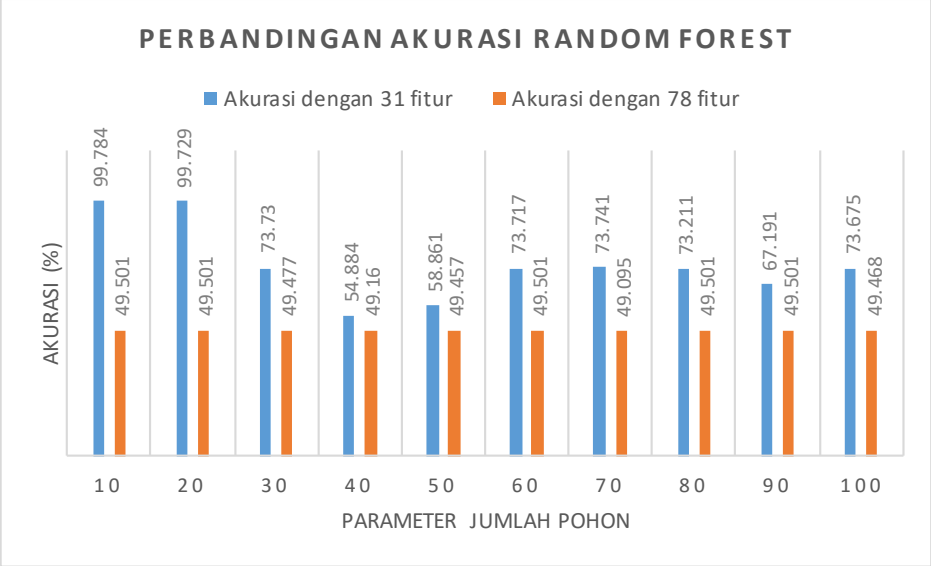
Pada skenario pengujian sistem ini menjelaskan mengenai simulasi serangan DDoS secara real time. Simulasi serangan dilakukan dengan menggunakan 2 proses utama. Adapun tahapan proses dari skenario pengujian dapat dilihat dari Gambar 2 dibawah ini. Pada proses pertama akan dilakukan pembuatan file log dengan dataset CSE-CIC-IDS2018. Hal ini dilakukan agar data packet serangan memiliki fitur yang sesuai dengan data training. Kemudian akan dilakukan injeksi data log agar data dapat digunakan pada proses kedua. Pada proses kedua, data log yang telah diinjeksi akan dibaca dan sistem akan mengklasifikasikan data log yang masuk kedalam kelas serangan atau normal menggunakan model klasifikasi random forest yang telah dibangun.

Gambar 2. Diagram Alir Proses Simulasi Serangan Real Time

3. Hasil dan Pembahasan

Pada bagian hasil dan pembahasan ini, dilakukan perbandingan terhadap jumlah fitur yang digunakan yaitu perbandingan antar 31 fitur dengan 78 fitur. Nilai parameter n_estimators atau jumlah pohon yang digunakan pada masing – masing model klasifikasi adalah n_estimators = [10, 20, 30, 40, 50, 60, 70, 80, 90, 100].

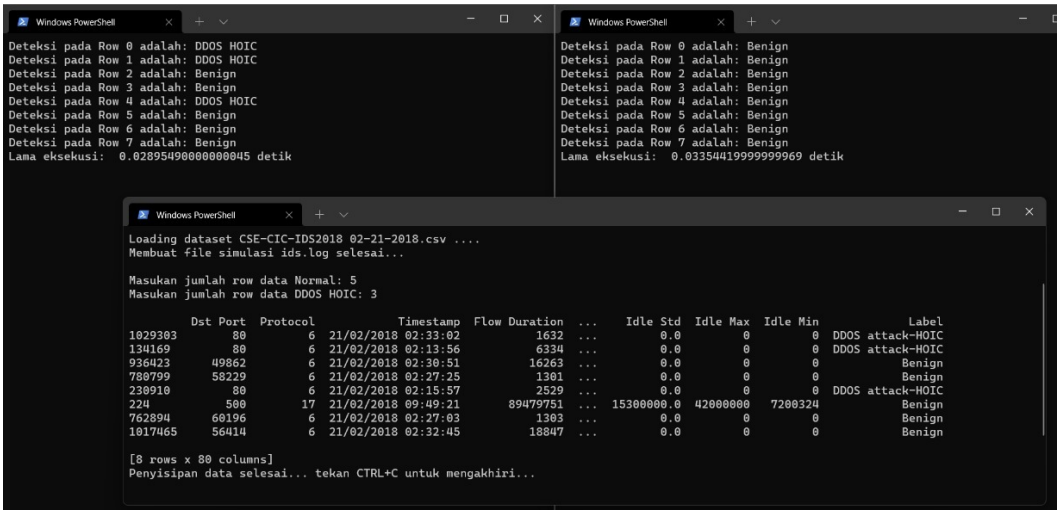
Pada Gambar 3 adalah grafik analisis perbandingan terhadap metode *Random Forest* dengan menggunakan 31 fitur dan 78 fitur.



Gambar 3. Hasil Akurasi

Pengujian dilakukan dengan menggunakan 2 jumlah fitur yang berbeda yaitu dengan 31 fitur dan 78 fitur. Jumlah 31 fitur didapatkan dengan memilih fitur yang memiliki nilai relasi diatas 0.1. Jumlah yang digunakan pada saat pengujian yaitu dengan rentang 10-100 pohon. Pada Gambar 3 dapat dilihat bahwa hasil pengujian tertinggi dihasilkan oleh nilai n_estimators = 10 pada kedua data. Pada data yang menggunakan 31 fitur hasil rata – rata akurasi tertinggi yaitu 99.784 % sedangkan rata – rata akurasi tertinggi pada data yang menggunakan 78 fitur adalah 49.501%.

Dari hasil pengujian tersebut dapat disimpulkan bahwa nilai parameter n_estimators atau jumlah pohon mempengaruhi hasil akurasi dari sistem. Nilai n_estimators optimal dari kedua data tersebut adalah 10. Selain itu, data dengan 31 fitur menghasilkan akurasi yang lebih tinggi dari pada data yang dengan 78 fitur. Namun dapat dilihat pada Gambar 3, pola yang dihasilkan dari pengujian sistem tidak sama.



Gambar 4. Perbandingan Lama Waktu Eksekusi Simulasi Real Time

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Berdasarkan waktu eksekusi yang dapat dilihat pada Gambar 4 dibutuhkan untuk melakukan testing dataset, diperoleh perbedaan waktu yang tidak berbeda jauh. Dimana waktu eksekusi sistem dengan model yang memiliki 31 fitur adalah 0.0289 detik, sedangkan waktu eksekusi sistem dengan model yang memiliki 78 fitur adalah 0.0335 detik. Dari kedua hasil tersebut, dapat dilihat bahwa perbedaan waktu eksekusi adalah kurang dari 0.005 detik yang berarti perbedaan waktu tidak terlalu besar.

4. Kesimpulan

Adapun kesimpulan yang diperoleh dari implementasi metode klasifikasi *Random Forest* dan seleksi fitur *Correlation-based Feature Selection* (CFS) untuk klasifikasi serangan *Distributed Denial of Service* (DDoS) dengan menggunakan dataset yang berasal dari CSE-CIC-IDS2018 adalah sebagai berikut. Setelah melakukan seleksi fitur terhadap dataset menggunakan *Correlation-based Feature Selection* (CFS), didapatkan hasil berupa nilai relasi antara fitur independen (X) dan fitur dependen (Y) untuk menentukan fitur-fitur yang relevan terhadap dataset. Seleksi fitur dilakukan dengan memilih fitur yang memiliki nilai relasi diatas 0.1 yang berjumlah 31 fitur. Untuk melakukan klasifikasi serangan *Distributed Denial of Service* (DDoS) menggunakan metode *Random Forest* dilakukan dengan mengambil *sample bootstrap* secara *random* dan membangun *tree* pada setiap *sample* yang diambil, kemudian pada proses terakhir dilakukan penentuan prediksi label dengan *majority vote* apakah log yang masuk tersebut termasuk ke dalam kelas data normal atau data serangan. Pada proses *training model*, dilakukan perubahan pada parameter n_estimators dengan rentang 10-100 untuk mendapatkan jumlah pohon yang paling baik. Pada penelitian ini didapatkan nilai n_estimators terbaik adalah 10. Akurasi yang dihasilkan oleh sistem dengan menggunakan metode *Random Forest* dan seleksi fitur *Correlation-based Feature Selection* (CFS) memiliki nilai rata-rata sebesar 99.784%. Akurasi tersebut diperoleh dari penggunaan parameter n_estimators yang bernilai 10. Sedangkan, pada model *Random Forest* tanpa seleksi fitur diperoleh akurasi tertinggi sebesar 49.501%. Hal ini menunjukkan tingginya akurasi dipengaruhi oleh perubahan pada parameter model *Random Forest* dan seleksi fitur *Correlation-based Feature Selection* (CFS).

Referensi

[1] D. B. Satmoko, P. Sukarno, and E. M. Jadied, “Peningkatan Akurasi Pendeteksian Serangan DDoS Menggunakan Multiclassifier Ensemble Learning dan Chi-Square,” *e-Proceeding Eng.*, vol. 5, no. 3, pp. 7977–7985, 2018.

[2] M. Aziz, R. Umar, and F. Ridho, “Implemetasi Jaringan Saraf Tiruan untuk Mendeteksi Serangan DDoS pada Forensik Jaringan,” *QUERY J. Sist. Inf.*, vol. 3, no. 1, pp. 46–52, 2019.

[3] E. L. Sofa and S. Subiyanto, “Routing Attacs Pada Internet of Things Berbasis Smart Intrution Detection System,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 2, pp. 329–338, 2020.

[4] K. Kurniabudi, A. Harris, and A. Rahim, “Seleksi Fitur dengan Information Gain Untuk Meningkatkan Deteksi Serangan DDoS Menggunakan Random Forest,” *Techno.Com*, vol. 19, no. 1, pp. 56–66, 2020.

[5] A. S. B. Asmoro, W. S. G. Irianto, and U. Pujianto, “Perbandingan Kinerja Hasil Seleksi Fitur pada Prediksi Kinerja Akademik Siswa Berbasis Pohon Keputusan,” *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, pp. 84–89, 2018.

[6] H. A. Yanti, H. Sukoco, and S. N. Neyman, “Pemodelan Identifikasi Trafik Bittorrent Dengan Pendekatan Correlation Based Feature Selection (CFS) Menggunakan Algoritme Decision Tree (C4.5),” *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 6, no. 1, pp. 1–9, 2021.

[7] S. Polamuri, “How the Random Forest Algorithm Works in Machine Learning,” 2017. http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/ (accessed Apr. 05, 2021).

[8] V. W. Siburian and I. E. Mulyana, “Prediksi Harga Ponsel Menggunakan Metode Random Forest,” *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 144–147, 2018.

This page is intentionally left blank