Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1470−1477
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6684

Komparasi Metode Random Forest Dan Support Vector Machine (SVM) Untuk Pemodelan Klasifikasi Serangan DDos

Christoper Michael Lauwl, Husain*, Baiq Nadila Nuzululnisa, HartonoWijaya

Fakultas Teknik, Ilmu Komputer, Universitas Bumigora, Mataram
Jl. Ismail Marzuki No.22, Cilinaya, Kec. Cakranegara, Kota Mataram, Nusa Tenggara Barat, Indonesia
Email: [1]24010820003@universitasbumigora.ac.id , [2]*husain@universitasbumigora.ac.id ,
[3]24010820002@universitasbumigora.ac.id, [4]24010820006@universitasbumigora.ac.id
Email Penulis Korespondensi: husain@universitasbumigora.ac.id
Submitted: **10/01/2025;** Accepted: **31/01/2025;** Published: **31/01/2025**

Abstrak−Serangan Distributed Denial of Service (DDoS) adalah jenis serangan siber yang bertujuan untuk membuat suatu layanan, jaringan, atau situs web menjadi tidak tersedia bagi pengguna sah. Tidak hanya hal tersebut DDOS menyebabkan crash pada server dengan cara melakukan pengiriman paket data secara berulang kali atau yang biasa disebut dengan spam. DDOS dapat diidentifikasi sebagai trafic anomali. Badan Siber dan Sandi Negara (BSSN) mencatat terdapat 403.990.813 trafic anomali dengan kasus 347 serangan berjenis DDOS. Berdasarkan permasalahan tersebut, diperlukan model yang dapat mengklasifikasikan serangan DDOS. Metode yang digunakan dalam penelitian ini yaitu Random Forest dan Support Vector Machine (SVM) dengan langkah yaitu data collection, loading dataset, preprocessing data, classification modeling, performance evaluating. Pada tahap akhir dari penelitian ini penulis mengambil metode terbaik atara kedua metode yang digunakan yaitu Random Forest dan Support Vector Machine. Hasil penelitian yang diperoleh yaitu Random Forest memiliki akurasi 99.9% sedangkan Support Vector Machine memiliki akurasi sebesar 97.0%, oleh karena itu dapat Random Forest memiliki akurasi lebih baik dalam melakukan klasifikasi serangan DDOS.

Kata Kunci: Random Forest; DDOS; Klasifikasi; Pemodelan; SVM;

Abstract−The Distributed Denial of Service (DDoS) attack is a type of cyberattack that aims to render a service, network, or website inaccessible to legitimate users. This attack not only disrupts services but also causes server crashes by repeatedly sending data packets, commonly referred to as spam. DDoS attacks can be identified as traffic anomalies. The National Cyber and Crypto Agency (BSSN) recorded 403,990,813 traffic anomalies with 347 cases specifically attributed to DDoS attacks. Based on this issue, a model capable of classifying DDoS attacks is necessary. This study employs the Random Forest and Support Vector Machine (SVM) methods through the steps of data collection, dataset loading, data preprocessing, classification modeling, and performance evaluation. In the final stage, the best method between Random Forest and Support Vector Machine is determined. The results indicate that Random Forest achieved an accuracy of 99.9%, whereas Support Vector Machine obtained an accuracy of 97.0%. Therefore, it can be concluded that Random Forest demonstrates better accuracy in classifying DDoS attacks.

Keywords: Random Forest; DDOS; Classification; Modelling; SVM;

1. PENDAHULUAN

Pada era sekarang, aspek keamanan dalam teknologi informasi adalah hal krusial yang perlu menjadi perhatian utama bagi setiap pengguna teknologi[1]. Koneksi jaringan global yang meluas di seluruh dunia memungkinkan untuk berkomunikasi antar pengguna dari berbagai wilayah, namun situasi ini juga berpotensi memunculkan berbagai ancaman terhadap sumber daya jaringan komputer para pengguna[2]. Salah satu ancaman tersebut adalah Distributed Denial of Service (DDOS). Serangan Distributed Denial of Service (DDOS) merupakan salah satu ancaman serius bagi situs web, dengan dampak signifikan yang dapat mengganggu keamanan sistem komputer[3].. Ancaman ini berpotensi membuat layanan, jaringan, atau situs web menjadi tidak dapat diakses oleh pengguna[4]. Selain itu, Serangan Distributed Denial of Service (DDOS) juga bisa menyebabkan server mengalami crash melalui pengiriman data yang terus-menerus, dengan cara kerja mengirimkan sejumlah besar perintah atau permintaan ke satu server dari berbagai komputer[5]. Serangan ini dikontrol oleh satu komputer atau pusat komando, sehingga server menjadi kewalahan dan akhirnya mengalami down atau gagal memberikan layanan[6]. Ketika server down, layanan tidak dapat diberikan kepada pengguna karena server telah menerima permintaan dalam jumlah berlebihan. Distributed Denial of Service (DDOS) dapat diidentifikasi sebagai trafic anomali. Menurut catatan Badan Siber dan Sandi Negara (BSSN), pada tahun 2023 terdapat 403.990.813 trafic anomali yang terdeteksi. Dari jumlah tersebut, tercatat sebanyak 347 insiden berupa serangan DDoS. Trafic anomali ini mengindikasikan aktivitas jaringan yang tidak biasa, yang bisa menjadi indikasi potensi ancaman bagi keamanan teknologi informasi. Jumlah serangan DDoS yang tercatat ini menyoroti kebutuhan untuk meningkatkan sistem pertahanan siber guna melindungi layanan penting dari gangguan yang merugikan (3). Metode yang digunakan dalam penelitian ini yaitu Random Forest dan Support Vector Machine (SVM).

Random Forest dan Support Vector Machine (SVM) adalah dua algoritma pembelajaran mesin yang sering digunakan dalam klasifikasi dan regresi, namun dengan pendekatan yang berbeda [7]. Random Forest merupakan metode ensemble yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi prediksi melalui metode voting mayoritas (pada klasifikasi) atau rata-rata (pada regresi), sehingga membuatnya lebih tahan terhadap overfitting dan cocok untuk dataset besar yang berisik[8]. Di sisi lain, SVM adalah

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1470−1477
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6684

algoritma berbasis margin yang mencari hyperplane optimal untuk memisahkan kelas-kelas dalam data, menggunakan kernel trick untuk menangani kasus non-linear dengan memetakan data ke dimensi yang lebih tinggi[9]. SVM memiliki keunggulan dalam menangani dataset yang lebih kecil dan terstruktur, tetapi sensitif terhadap outlier dan memerlukan tuning parameter yang tepat untuk performa optimal[10]. Meskipun berbeda, kedua algoritma ini dapat menghasilkan model prediksi yang kuat dalam berbagai aplikasi, bergantung pada karakteristik dan kebutuhan dataset. Dalam penelitian yang dilakukan oleh penulis memiliki 5 tahap yaitu data collection , load data, pre-processing data, melakuka pemodelan dengan RF, dan SVM, memilih akurasi terbaik dari kedua model tersebut.

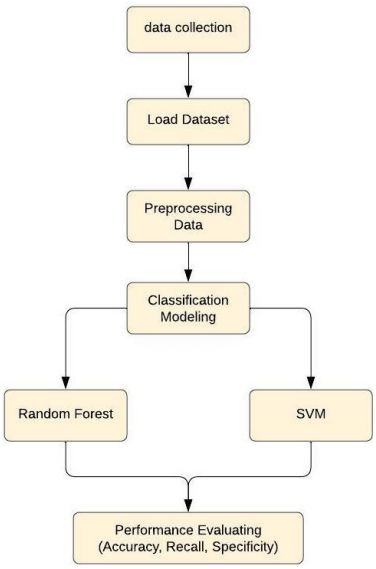
Adapun penelitian sebelumnya yang dilakukan oleh penelitian tersebut dilakukan untuk pembuatan model klasifikasi dan prediksi untuk serangan DDOS, penelitian tersebut menggunakan 5 metode yaitu convolutional neural network (CNN) , long short term memory (LSTM), support vector machine (SVM) , Random Forest Hasil dari penelitian tersebut yaitu CNN dan LSTOm memiliki akurasi sebesar 85%, SVM memiliki akurasi sebesar 78.34%, RF menghasilkan akurasi sebesar89%, XGBost menghasilkan akurasi sebesar 90%[11]. Penelitian selanjutnya dilakukan oleh dengan judul Klasifikasi Serangan Distributed Denial-of-Service (DDoS) menggunakan Metode Data Mining Naïve Bayes dengan 6 atribut data yang digunakan dengan 1000 jumlah data, dengan akurasi sebesar 95%. Penelitian yang dilakukan oleh yang berjudul “DDoS Attack Classification on Cloud Environment Using Machine Learning Techniques with Different Feature Selection Methods” dengan menggunakan 8 atribut data dan 26.283 jumlah data dengan menggunakan metode Naive Bayes yang menghasilkan akurasi sebesar 91%, Decission Tree sebesar 92%, dan SVM sebesar 98% [12]. Terdapat pula penelitian yang dilakukan oleh yang menggunakan metode k-Nearest Neighbor (KNN) dengan akurasi 95%, Decision Tree (DT) dengan akurasi 94%, Artificial Neural Network (ANN) dengan akurasi dengan akurasi 97%, Support Vector Machine (SVM) dengan akurasi 80% [13]. Terakhir penelitian dengan judul “SSK-DDoS: distributed stream processing framework based classification system for DDoS attacks”dengan menggunakan metode yang sama, dengan hasil akurasi sebesar 80% [14]. Oleh karena itu, pada penelitian ini dilakukan penambahan atribut dataset, dan perbantingan antara SVM dan Random Forest untuk memperoleh kualitas akurasi yang lebih tinggi dari penelitian sebelumnya.

Dari hasil penelitian terdahulu, dapat menunjukkan bahwa algoritma seperti Random Forest, Gradient Boosting, dan metode deep learning seperti CNN dan LSTM telah digunakan untuk mendeteksi serangan DDoS dengan hasil akurasi tinggi. Namun, perbandingan langsung antara Random Forest dan SVM untuk klasifikasi DDoS masih jarang dilakukan. Penelitian semacam ini perlu dilakukan karena dapat untuk mengevaluasi dari kinerja dari suatu metode untuk melihat keefektifan dari kedua algoritma yang dipakai yaitu random forest dan SVM dan diharapkan dapat memberikan hasil dari analisis kinerja dan akurasi yang tinggi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada Gambar 2.1 merupakan tahapan penelitian yang digunakan sebelum melakukan pemodelan klasifikasi serangan. Terdapat 7 langkah yang dimulai dari pengumpulan data, loading data, preprocessing data, dan dilakukan pemodelan dengan RF dan SVM.



Gambar 1 Kerangka Penelitian

Berikut ini merupakan penjelasan langkah-langkah yang ada di Gambar 1. Kerangka Penelitian yaitu:

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1470−1477
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6684

a) Data Collection
Tahap Data Collection adalah langkah pertama dalam proses analisis dan pemodelan data, di mana data yang akan digunakan dalam penelitian dikumpulkan [15]. Pada penelitian ini, data dikumpulkan dari situs Kaggle.com, yang merupakan platform populer untuk berbagi dan mengakses dataset publik dalam berbagai domain [16]. Dataset yang digunakan berkaitan dengan lalu lintas jaringan untuk mendeteksi serangan Distributed Denial of Service (DDoS), dengan total jumlah data sebanyak 104.345 baris dan terdiri dari 23 atribut [17]. Atribut-atribut ini mencakup informasi penting seperti alamat IP sumber (source IP), alamat IP tujuan (destination IP), jumlah paket (packet count), jumlah byte data (byte count), durasi (duration), serta label (indikasi apakah koneksi termasuk anomali atau tidak). Mengumpulkan data dari sumber yang andal seperti Kaggle penting untuk memastikan bahwa dataset yang digunakan cukup kaya dan representatif, sehingga model yang dilatih mampu mengenali pola lalu lintas jaringan yang normal dan anomali secara akurat.

b) Load Dataset
Setelah data dikumpulkan, langkah berikutnya adalah memuat dataset ke dalam lingkungan pemrograman, sehingga data tersebut dapat diakses dan diproses lebih lanjut. Dataset DDoS yang diperoleh dari Kaggle ini biasanya dibaca ke dalam bentuk DataFrame menggunakan pustaka pandas, yang sangat populer dalam pemrosesan data di Python. DataFrame memungkinkan data diorganisir dalam bentuk tabel, sehingga memudahkan dalam eksplorasi, pemfilteran, dan manipulasi. Memuat dataset ini adalah langkah esensial agar data tersedia dalam format yang dapat dipahami dan digunakan oleh algoritma machine learning. Dengan dataset dalam bentuk DataFrame, kita dapat melakukan analisis awal, seperti melihat beberapa baris pertama, mengidentifikasi tipe data untuk setiap atribut, dan memeriksa konsistensi serta kelengkapan data. Langkah ini merupakan persiapan penting sebelum masuk ke tahap pra-pemrosesan dan pemodelan data.

c) Preprocessing Data
Tahap Preprocessing Data adalah salah satu bagian penting dalam pipeline machine learning, yang bertujuan untuk membersihkan dan menyiapkan data agar siap digunakan oleh model. Dalam tahap ini, dilakukan beberapa langkah yang memastikan bahwa data yang akan dimasukkan ke dalam model memiliki kualitas tinggi, bebas dari inkonsistensi, dan dalam bentuk yang mudah diproses oleh algoritma. Langkah pertama dalam pra-pemrosesan adalah mengisi nilai yang hilang atau menangani data yang kosong (missing values), misalnya dengan metode imputer atau menggantinya dengan nilai rata-rata (untuk data numerik) atau modus (untuk data kategori). Selanjutnya, fitur numerik dalam dataset mungkin perlu dinormalisasi atau distandarisasi untuk memastikan bahwa skala data tidak mengganggu kinerja model. Selain itu, fitur kategori mungkin perlu di-encode atau diubah menjadi format numerik, misalnya dengan teknik one-hot encoding. Terakhir, proses ini juga mencakup identifikasi dan penanganan outliers atau nilai pencilan yang berpotensi mempengaruhi prediksi model. Tahap ini dirancang untuk memaksimalkan kualitas data sehingga model machine learning dapat menghasilkan hasil yang lebih akurat dan andal.

d) Classification Modelling
Setelah data diproses, langkah berikutnya adalah membangun model untuk klasifikasi menggunakan dua algoritma berbeda, yaitu Random Forest dan Support Vector Machine (SVM). Kedua model ini dipilih karena memiliki pendekatan yang berbeda dan masing-masing memiliki keunggulan khusus. Pada tahap ini, dataset yang sudah dipra-pemrosesan digunakan untuk melatih kedua model, yang kemudian dievaluasi berdasarkan kinerja masing-masing. Untuk membandingkan kinerja, metrik akurasi digunakan, di mana model dengan akurasi tertinggi akan dipilih sebagai model terbaik. Dengan membandingkan hasil dari Random Forest dan SVM, kita dapat menentukan model mana yang lebih cocok untuk mendeteksi serangan DDoS dalam dataset ini. Langkah ini bertujuan untuk memastikan bahwa hasil klasifikasi dapat diandalkan dan berguna dalam skenario aplikasi nyata.

e) Random Forest
Random Forest adalah algoritma pembelajaran mesin berbasis ensemble yang menggunakan sekumpulan pohon keputusan untuk meningkatkan akurasi dan stabilitas prediksi. Algoritma ini dikembangkan oleh Leo Breiman dan Adele Cutler sebagai pengembangan dari metode pohon keputusan tunggal yang cenderung rentan terhadap overfitting. Dalam Random Forest, setiap pohon dilatih menggunakan subset acak dari data, dan hasil prediksi diperoleh dengan cara voting mayoritas (untuk klasifikasi) atau rata-rata (untuk regresi). Dengan menggabungkan banyak pohon, Random Forest mampu menangani data yang berisik dan memiliki performa yang baik pada data dengan banyak fitur. Algoritma ini juga dapat memberikan pentingnya setiap fitur, yang berguna untuk interpretasi hasil[18][19][20].

f) Support Vector Machine (SVM)
Support Vector Machine (SVM) adalah algoritma klasifikasi yang bekerja dengan mencari hyperplane optimal yang memisahkan kelas-kelas data dalam ruang fitur. SVM berfokus pada penciptaan margin maksimum antara kelas yang berbeda, yang berarti bahwa SVM tidak hanya memperhatikan pemisahan data tetapi juga memaksimalkan jarak antara kelas untuk meningkatkan akurasi dan generalisasi. Dalam kasus data yang tidak dapat dipisahkan secara linear, SVM menggunakan kernel trick untuk memetakan data ke ruang dimensi yang lebih tinggi, sehingga memungkinkan pemisahan. Meskipun SVM efektif dalam

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1470−1477
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6684

menangani dataset yang lebih kecil dan terstruktur, algoritma ini memerlukan tuning parameter yang tepat untuk performa optimal, terutama dalam memilih kernel yang sesuai[21]–[25].

3. HASIL DAN PEMBAHASAN

3.1 Data Collection

Tahap pengumpulan data, dilakukan melalui situs kaggle.com, dimana diperolah data dengan jumlah fitur sebanyak 25. Pada tabel 1. merupakan atribut data serangan DDOS yang terdapat pada dataset.

Tabel 1. Atribut Data Serangan DDOS

Atribut	Jenis	Keterangan
Dt	Int	Catatan waktu lalu lintas data
Switch	Int	ID perangkat switch transmisi data
Src	Object	Alamat IP sumber
Dst	Object	Alamat IP tujuan
pktcount	Int	Jumlah paket yang dikirimkan dalam sesi atau periode waktu tertentu
bytecount	Int	Jumlah byte yang dikirimkan dalam sesi. Mengukur total ukuran data dalam byte.
Dur	Int	Durasi sesi atau koneksi dalam satuan waktu (misalnya, milidetik).
Dur_nsec	Int	Durasi dalam nanodetik, yang lebih akurat untuk melacak waktu hingga nanodetik.
Tot_dur	Float	Total durasi dari sesi atau aliran data dalam jaringan.
Flows	Int	Jumlah aliran (flows) data dalam periode tertentu.
Packettins	Int	Jumlah paket packet_in yang dihasilkan oleh switch ke pengontrol.
pktperflow	Int	Jumlah rata-rata paket per aliran data (flow)
byteperflow	Int	Jumlah rata-rata byte per aliran data.
Pktrate	Int	Kecepatan pengiriman paket per detik.
Pairflow	Int	Mencatat apakah aliran memiliki flow berpasangan (balikan) dari src ke dst dan sebaliknya. Bisa menunjukkan komunikasi dua arah.
Protocol	object	Protokol yang digunakan dalam komunikasi, misalnya TCP, UDP, ICMP,
port_no	Int	Nomor port yang terlibat dalam koneksi jaringan
tx_bytes	Int	Total byte yang ditransmisikan (dikirim) oleh perangkat selama sesi atau aliran data.
rx_bytes	Int	Total byte yang diterima oleh perangkat selama sesi atau aliran data.
tx_kbps	Int	Kecepatan transmisi data dalam kilobit per detik (kbps).
rx_bytes	Int	Kecepatan penerimaan data dalam kilobit per detik.
tx_kbps	Int	Kecepatan transmisi data dalam kilobit per detik (kbps). Mengukur kecepatan unggah atau pengiriman data.
rx_kbps	Float	Kecepatan penerimaan data dalam kilobit per detik. Mengukur kecepatan unduh atau penerimaan data.
tot_kbps	Float	Total kecepatan lalu lintas data (gabungan tx_kbps dan rx_kbps). Menunjukkan total bandwidth yang digunakan.
Label	Int	Label target yang mungkin menunjukkan apakah data tersebut normal atau termasuk serangan (misalnya, 0 untuk normal dan 1 untuk serangan).

3.2 Load Dataset

Pada tahap Load Dataset, data mentah dari sistem jaringan diimpor ke dalam lingkungan kerja untuk dianalisis lebih lanjut. Data ini terdiri dari berbagai atribut yang merepresentasikan karakteristik lalu lintas jaringan, seperti dt (timestamp data), switch (ID dari perangkat switch yang menangani lalu lintas), src dan dst (alamat IP sumber dan tujuan), pktcount (jumlah paket yang ditransmisikan), bytecount (jumlah byte data), dur (durasi koneksi dalam detik), tot_kbps (kecepatan total dalam kilobit per detik), dan label (indikator yang menyatakan apakah data tersebut tergolong anomali atau tidak). Contoh data yang dihasilkan menunjukkan beberapa baris pertama, di mana setiap baris mewakili satu koneksi atau aliran data di jaringan. Dengan memuat dataset ini, pengguna dapat mulai melakukan eksplorasi data untuk memahami pola lalu lintas jaringan serta mengidentifikasi perilaku normal dan anomali sebagai persiapan untuk tahap analisis selanjutnya.

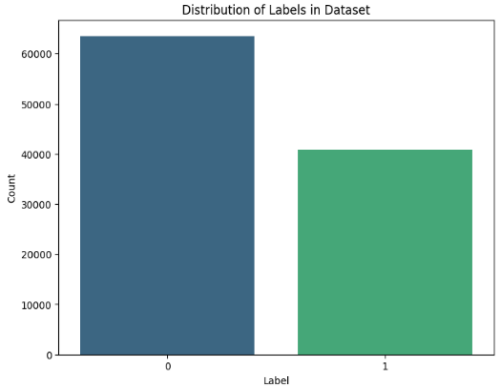
Tabel 2. Hasil dari Load Dataset

dt	switch	src	dst	pktcount	…	bytecount	dur	tot_kbps	label
0	11425	1	10.0.0.1	10.0.0.8	…	45304	48294064	0.0	0.0
1	11605	1	10.0.0.1	10.0.0.8	…	126395	1,35E+08	0.0	0.0
2	11425	1	10.0.0.2	10.0.0.8	…	90333	96294978	0.0	0.0
3	11425	1	10.0.0.2	10.0.0.8	…	90333	96294978	0.0	0.0
4	11425	1	10.0.0.2	10.0.0.8	…	90333	96294978	0.0	0.0

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1470−1477
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6684

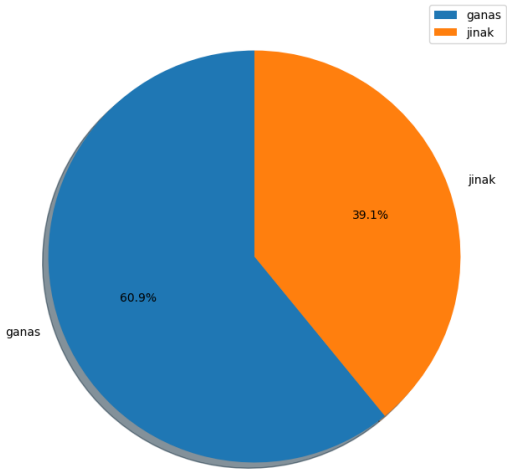
3.3 Preprocessing Data

Dalam pre-processing data pada penelitian ini, tahap pertama yang dilakukan yaitu melihat data serangan dengan identitas 0 yaitu diklasifikasikan sebagai serangan yang tidak berbahaya atau jinak dan 1 diklasifikasikan sebagai yang berbahaya atau ganas. Grafik distribusi dari kedua data tersebut dapat terhilat pada gambar 3.1 berikut:



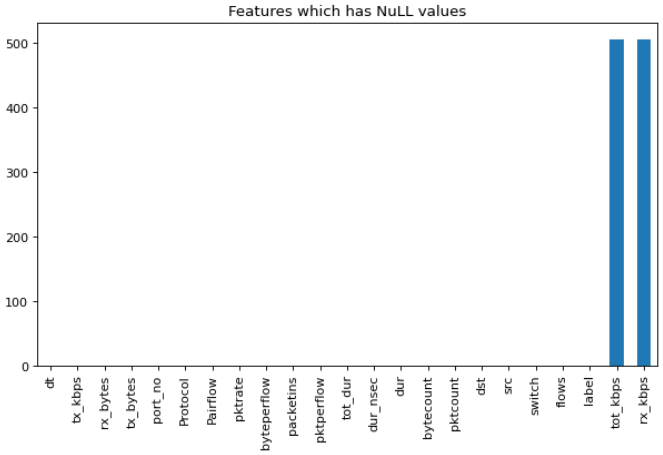
Gambar 2. Grafik Distrbusi

Pada Gambar 2. terlihat distribusi data antara serangan DDoS yang "ganas" dan yang "jinak". Grafik pie menunjukkan bahwa proporsi data yang masuk dalam kategori "ganas" lebih besar, yaitu 60,9%, sedangkan kategori "jinak" memiliki persentase sebesar 39,1%. Hal ini menunjukkan bahwa sebagian besar data tergolong dalam serangan yang lebih berbahaya atau mengakibatkan dampak signifikan, sementara sisanya merupakan serangan yang lebih ringan atau memiliki efek yang lebih terbatas.



Gambar 3. Grafik Pie

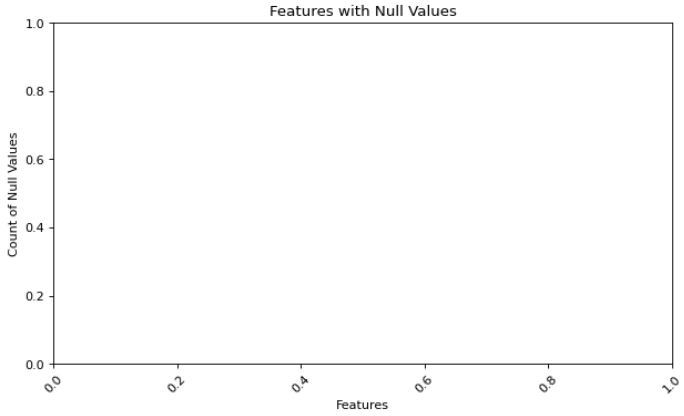
Pada tahap selanjutnya dalam pre-processing data dilakukan penghapusan data kosong, dimana terdapat sekitar 506 data kosong pada fitur rx_kbps dan tot_kbps dengan grafik nilai kosong sebagai berikut.



Gambar 4. Grafik Nilai Kosong 2 Fitur

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1470−1477
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6684

Pada Gambar 4. diatas terlihat dalam 2 fitur tersebut masih terdapat nilai kosong, oleh karena itu dilakukan penghapusan nilai kosong dengan hasil grafik nilai kosong seperti yang terlihat pada gambar dibawah ini.



Gambar 5. Grafik Nilai Kosong

Pada Gambar 5. merupakan grafik data yang telah di bersihkan, yang pada tahap sebelumnya dilakukan pemeriksaan data, masih terdapat data yang kosong atau bernilai null, dalam grafik diatas terlihat bahwa sudah tidak terdapat lagi data yang bernilai kosong. Pada tahap seleksi fitur penulis melakukan seleksi untuk fitur yang saling keterkaitan, dalam penelitian ini terdapat 10 fitur yang saling berkaitan antara lain yaitu 'dt', 'src', 'pktcount', 'bytecount', 'dur', 'tot_dur', 'flows', 'pktperflow', 'pktrate', 'Protocol'. Adapun penjelasan fitur yang digunakan terdapat pada tabel dibawah.

Tabel 3. Hasil Tahap Seleksi Fitur

Atribut	Jenis	Keterangan
Dt	Int	Catatan waktu lalu lintas data
Src	Object	Alamat IP sumber
pktcount	Int	Jumlah paket yang dikirimkan dalam sesi atau periode waktu tertentu
bytecount	Int	Jumlah byte yang dikirimkan dalam sesi. Mengukur total ukuran data dalam byte.
Dur	Int	Durasi sesi atau koneksi dalam satuan waktu (misalnya, milidetik).
Tot_dur	Float	Total durasi dari sesi atau aliran data dalam jaringan.
Flows	Int	Jumlah aliran (flows) data dalam periode tertentu.
pktperflow	Int	Jumlah rata-rata paket per aliran data (flow)
Pktrate	Int	Kecepatan pengiriman paket per detik.
Protocol	object	Protokol yang digunakan dalam komunikasi, misalnya TCP, UDP, ICMP,

3.4 Classification Modelling

Hasil dari model klasifikasi berdasarkan 2 metode diatas yang digunakan yaitu Random Forest (RF) dan Naive Bayes yaitu dapat terlihat pada tabel dibawah ini. Hasil dari Algoritma Support Vector Machie dapat terlihat pada tabel 4. berikut

Tabel 4. Hasil Algoritma SVM

Class	Precision	Recall	F1-Score	Support
0	0.90	0.96	0.93	17757
1	0.95	0.86	0.90	13395
Accuracy			0.92	31152
Macro Avg	0.92	0.91	0.91	31152
Weighted Avg	0.92	0.92	0.92	31152

Pada Tabel 4. menunjukkan hasil evaluasi kinerja model klasifikasi yang menggunakan beberapa metrik: Precision, Recall, F1-Score, dan Support. Berdasarkan hasilnya, model menunjukkan performa yang baik pada kedua kelas (kelas 0 dan kelas 1). Untuk kelas 0, Precision adalah 0.90, yang berarti 90% dari prediksi positif untuk kelas ini adalah benar. Sementara itu, Recall untuk kelas 0 mencapai 0.96, yang menunjukkan bahwa 96% dari semua data kelas 0 berhasil diidentifikasi dengan benar oleh model. F1-Score untuk kelas 0 adalah 0.93, yang menunjukkan keseimbangan yang baik antara Precision dan Recall. Untuk kelas 1, Precision adalah 0.95, yang berarti 95% dari prediksi positif untuk kelas ini akurat. Namun, Recall untuk kelas 1 sedikit lebih rendah, yaitu 0.86, menunjukkan bahwa 86% dari data kelas 1 teridentifikasi dengan benar. F1-Score untuk kelas 1 adalah 0.90, mencerminkan keseimbangan yang baik meskipun Recall sedikit lebih rendah dibandingkan kelas 0.

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1470−1477
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6684

Secara keseluruhan, accuracy model adalah 92%, yang berarti model berhasil mengklasifikasikan 92% data dengan benar. Macro Average menunjukkan rata-rata dari Precision, Recall, dan F1-Score tanpa memperhitungkan distribusi kelas, yang masing-masing bernilai 0.92, 0.91, dan 0.91. Sementara itu, Weighted Average memperhitungkan distribusi kelas dan memberikan nilai 0.92 untuk Precision, 0.92 untuk Recall, dan 0.92 untuk F1-Score, menunjukkan bahwa model bekerja dengan baik meskipun ada perbedaan jumlah data antara kelas 0 dan kelas 1. Dengan demikian, model menunjukkan performa yang sangat baik dalam mengenali dan mengklasifikasikan data dari kedua kelas secara keseluruhan.

Tabel 5. Hasil Evaluasi Model Random Forest

Class	Precision	Recall	F1-Score	Support
0	0.99	1.00	1.00	18922
1	1.00	0.99	0.99	12230
Accuracy			0.99	31152
Macro Avg	0.99	0.99	0.99	31152
Weighted Avg	0.99	0.99	0.99	31152

Tabel 5. merupakan hasil evaluasi model Random Forest (RF) menunjukkan performa yang sangat mengesankan dengan accuracy mencapai 99.42%, artinya model ini berhasil mengklasifikasikan hampir 99.5% data dengan benar. Untuk kelas 0, model memiliki Precision sebesar 0.99, yang menunjukkan bahwa 99% dari prediksi untuk kelas ini adalah akurat, dengan Recall sempurna 1.00, yang berarti semua data kelas 0 berhasil dikenali dengan benar. F1-Score untuk kelas 0 mencapai 1.00, menandakan keseimbangan yang sempurna antara Precision dan Recall. Sedangkan untuk kelas 1, Precision juga sangat tinggi, yaitu 1.00, menunjukkan bahwa seluruh prediksi positif untuk kelas ini akurat, meskipun Recall sedikit lebih rendah di angka 0.99, yang berarti 99% dari data kelas 1 berhasil dikenali. F1-Score untuk kelas 1 adalah 0.99, yang tetap menunjukkan kinerja yang sangat baik meskipun ada sedikit perbedaan dengan kelas 0. Secara keseluruhan, model ini memiliki accuracy 99%, yang menunjukkan kinerja yang sangat luar biasa. Nilai Macro Average untuk Precision, Recall, dan F1-Score masing-masing mencapai 0.99, mencerminkan bahwa model bekerja sangat baik meskipun ada ketidakseimbangan jumlah data antar kelas. Begitu juga dengan Weighted Average, yang menghasilkan nilai 0.99 untuk semua metrik, menunjukkan bahwa model tetap efektif meskipun kelas 0 memiliki jumlah data yang lebih besar. Secara keseluruhan, Random Forest berhasil memberikan klasifikasi yang sangat akurat dan seimbang antara kedua kelas dalam dataset.

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan, diperoleh bahwa metode Random Forest dan Support Vector Machine (SVM) memiliki keunggulan masing-masing dalam klasifikasi serangan Distributed Denial of Service (DDoS). Random Forest menunjukkan performa yang lebih baik pada data yang lebih besar, sedangkan SVM lebih efektif untuk data terstruktur dan lebih kecil. Algoritma Ranfom Forest memiliki akurasi yang jauh lebih tinggi dari SVM, dengan akurasi sebesar 99.9% sedangkan SVM memiliki akurasi sebesar Data dari Badan Siber dan Sandi Negara (BSSN) menunjukkan tingginya jumlah trafik anomali yang terkait dengan serangan DDoS, mengindikasikan perlunya sistem deteksi yang andal. Penelitian ini mengkonfirmasi bahwa model berbasis Random Forest dan SVM dapat menjadi alat efektif dalam deteksi DDoS. Di sisi lain, hasil penelitian terdahulu mengindikasikan bahwa metode deep learning seperti CNN dan LSTM juga menunjukkan akurasi yang tinggi dalam deteksi serangan DDoS. Dengan demikian, pemodelan klasifikasi berbasis Random Forest dan SVM dapat menjadi solusi yang kuat dalam mendeteksi serangan DDoS, namun model deep learning memiliki potensi lebih tinggi dalam menghadapi pola serangan yang kompleks.

REFERENCES

[1] N. S. Dinarti, S. R. Salsabila, dan Y. T. Herlambang, “Dilema Etika dan Moral dalam Era Digital: Pendekatan Aksiologi Teknologi terhadap Privasi Keamanan, dan Kejahatan Siber,” Daya Nas. J. Pendidik. Ilmu-Ilmu Sos. dan Hum., vol. 2, no. 1, hal. 8–16, 2024.
[2] F. Indah, A. Q. Sidabutar, dan N. A. Nasution, “Peran cyber security terhadap keamanan data penduduk negara Indonesia (Studi kasus: Hacker Bjorka),” J. Bid. Penelit. Inform., vol. 1, no. 1, hal. 57–64, 2023.
[3] M. J. Alhafiz, A. Fauzi, A. Dwiansyah, B. R. Indriani, F. M. A. Putra, dan R. R. Ridwani, “Dampak Denial of Service pada Perusahaan Perbankan di Indonesia,” J. Ilmu Multidisplin, vol. 2, no. 1, hal. 114–120, 2023.
[4] R. D. Hapsari dan K. G. Pambayun, “Ancaman cybercrime di indonesia: Sebuah tinjauan pustaka sistematis,” J. Konstituen, vol. 5, no. 1, hal. 1–17, 2023.
[5] M. Ikum, “Implementasi Packet Tracer 8.0 Pada Simulator Pintu Rumah Pintar Berbasis Teknologi Radio Frequency Identification.” Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta, 2024.
[6] S. Muhammad Rifqi Noval, “HUKUM SIBER Kebangkitan Kembali Metaverse Beserta Permasalahan Hukumnya,” 2024.
[7] P. K. Sari dan R. R. Suryono, “Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse,” J. Mnemon., vol. 7, no. 1, hal. 31–39, 2024.

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1470−1477
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6684

[8] A. F. Nugraha, R. F. A. Aziza, dan Y. Pristyanto, "Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing," J. Infomedia Tek. Inform. Multimed. Jar., vol. 7, no. 1, 2022.
[9] A. Widiyanti dan D. A. Megawaty, "Perbandingan Algoritma K-Nearest Neighbor dan Support Vector Machine Pada Pengenalan Pola Tulisan Tangan," J. MEDIA Inform. BUDIDARMA, vol. 8, no. 3, hal. 1451–1459, 2024.
[10] B. A. Maulana, M. J. Fahmi, A. M. Imran, dan N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM): Sentiment Analysis of Pluang Applications With Naive Bayes and Support Vector Machine (SVM) Algorithm," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 2, hal. 375–384, 2024.
[11] Ismail et al., "A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks," IEEE Access, vol. 10, hal. 21443–21454, 2022, doi: 10.1109/ACCESS.2022.3152577.
[12] C. Bagyalakshmi, "DDoS Attack Classification on Cloud Environment Using Machine Learning Techniques with Different Feature Selection Methods," Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 5, hal. 7301–7308, 2020, doi: 10.30534/ijatcse/2020/60952020.
[13] Ö. Tonkal, H. Polat, E. Başaran, Z. Cömert, dan R. Kocaoğlu, "Machine learning approach equipped with neighbourhood component analysis for ddos attack detection in software-defined networking," Electron., vol. 10, no. 11, 2021, doi: 10.3390/electronics10111227.
[14] N. V. Patil, C. R. Krishna, dan K. Kumar, "SSK-DDoS: distributed stream processing framework based classification system for DDoS attacks," Cluster Comput., vol. 25, no. 2, hal. 1355–1372, 2022, doi: 10.1007/s10586-022-03538-x.
[15] N. Winarti, L. H. Maula, A. R. Amalia, dan N. L. A. Pratiwi, "Penerapan Model Pembelajaran Project Based Learning Untuk Meningkatkan Kemampuan Berpikir Kritis Siswa Kelas III Sekolah Dasar," J. Cakrawala Pendas, vol. 8, no. 3, hal. 552–563, 2022.
[16] D. P. MAWARDI, "DETEKSI AWAL KLASIFIKASI JENIS PENYAKIT KANKER KULIT DENGAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK (CNN) BERBASIS MOBILE APPS." UNIVERSITAS PGRI SEMARANG, 2024.
[17] Y. Cao, Y. Kang, C. Wang, dan L. Sun, "Instruction Mining: When Data Mining Meets Large Language Model Finetuning," arXiv Prepr. arXiv2307.06290, 2023.
[18] F. A. Larasati, D. E. Ratnawati, dan B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," J. Pengemb. Teknol. Inf. Dan Ilmu Komput., vol. 6, no. 9, hal. 4305–4313, 2022.
[19] M. M. Mutoffar dan A. Fadillah, "Klasifikasi Kualitas Air Sumur Menggunakan Algoritma Random Forest," Naratif J. Nas. Riset, Apl. dan Tek. Inform., vol. 4, no. 2, hal. 138–146, 2022, doi: 10.53580/naratif.v4i2.160.
[20] H. Tantyoko, D. K. Sari, dan A. R. Wijaya, "Prediksi potensial gempa bumi Indonesia menggunakan metode random forest dan feature selection," IDEALIS Indones. J. Inf. Syst., vol. 6, no. 2, hal. 83–89, 2023.
[21] I. S. K. Idris, Y. A. Mustofa, dan I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," Jambura J. Electr. Electron. Eng., vol. 5, no. 1, hal. 32–35, 2023.
[22] H. S. Wafa, A. I. Hadiana, dan F. R. Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," vol, vol. 4, hal. 40–45, 2022.
[23] E. Suryati, S. Styawati, dan A. A. Aldino, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," J. Teknol. dan Sist. Inf., vol. 4, no. 1, hal. 96–106, 2023.
[24] H. Harnelia, "Analisis Sentimen Review Skincare Skintific Dengan Algoritma Support Vector Machine (Svm)," J. Inform. dan Tek. Elektro Terap., vol. 12, no. 2, 2024.
[25] D. Oktavia, Y. R. Ramadahan, dan M. Minarto, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," KLIK Kaji. Ilm. Inform. dan Komput., vol. 4, no. 1, hal. 407–417, 2023.