RESEARCH ARTICLE • OPEN ACCESS

Analisis Komparatif *Random Forest* dan *Support Vector Machine* untuk Klasifikasi Tingkat Keparahan Serangan Siber

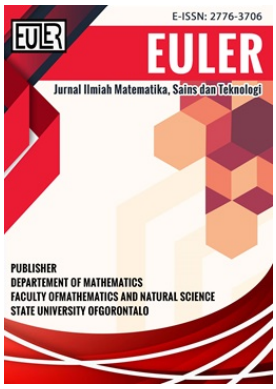
Reyhanssan Islamey, Sri Winiarti, dan Imam Riadi

Volume 14, Issue 1, Pages 1–14, April 2026

Diterima 4 Desember 2025, Direvisi 17 Februari 2026, Disetujui 23 Februari 2026, Diterbitkan 6 April 2026

To Cite this Article : R. Islamey, S. Winiarti, dan I. Riadi,"Analisis Komparatif *Random Forest* dan *Support Vector Machine* untuk Klasifikasi Tingkat Keparahan Serangan Siber", *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 14, no. 1, pp. 1–14, 2026, *https://doi.org/10.37905/euler.v14i1.36558*

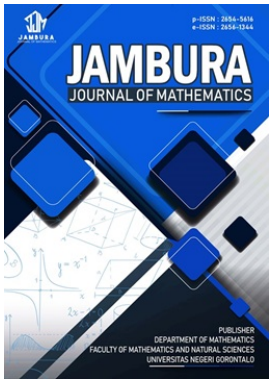
JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI



Homepage	:	http://ejurnal.ung.ac.id/index.php/euler/index
Journal Abbreviation	:	Euler J. Ilm. Mat. Sains dan Teknol.
Frequency	:	Three times a year
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.37905/euler
Online ISSN	:	2776-3706
License	:	Creative Commons Attribution-NonCommercial 4.0 International License
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/euler/oai
Google Scholar ID	:	QF_r-_gAAAAJ
Email	:	euler@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

Jambura Journal of Biomathematics



Jambura Journal of Mathematics

Jambura Journal of Mathematics Education

Jambura Journal of Probability and Statistics

Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi, Volume 14, Issue 1, Pages 1–14, April 2026 *https://doi.org/10.37905/euler.v14i1.36558*

Research Article

Check for updates

Analisis Komparatif *Random Forest* dan *Support Vector Machine* untuk Klasifikasi Tingkat Keparahan Serangan Siber

Reyhanssan Islamey[1], **Sri Winiarti**[1,*], **Imam Riadi**[1]

[1]*Program Studi Informatika, Universitas Ahmad Dahlan, Yogyakarta 55166, Indonesia*

ARTICLE HISTORY
Diterima 4 Desember 2025
Direvisi 17 Februari 2026
Disetujui 23 Februari 2026
Diterbitkan 6 April 2026

KATA KUNCI
Keamanan Siber
Klasifikasi Keparahan Serangan
Machine Learning
Random Forest
Support Vector Machine

KEYWORDS
Cyber Security
Attack Classification
Machine Learning
Random Forest
Support Vector Machine

ABSTRAK. *Meningkatnya volume dan kompleksitas serangan siber pada infrastruktur jaringan yang memproses jutaan lalu lintas data setiap hari menyebabkan tim keamanan kewalahan dalam menentukan prioritas respons secara cepat dan akurat, sebuah fenomena yang dikenal sebagai alert fatigue. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Support Vector Machine (SVM) dan Random Forest (RF) dalam klasifikasi tingkat keparahan serangan siber (Low, Medium, High). Studi ini menggunakan dataset publik Cyber Security Attacks yang terdiri dari 40.000 rekaman lalu lintas jaringan dan direduksi menjadi 20.000 data bersih melalui pra-pemrosesan dan rekayasa fitur. Metodologi mencakup pembersihan data, seleksi 10 fitur signifikan menggunakan SelectKBest, standarisasi fitur numerik, serta evaluasi model melalui tiga skenario pembagian data (70:30, 80:20, dan 90:10) dengan pendekatan stratified splitting. Hasil eksperimen menunjukkan bahwa SVM secara konsisten mengungguli RF pada seluruh skenario, dengan performa terbaik pada skenario 80:20, yaitu akurasi sebesar 98,92% dan F1-Score (weighted average) sebesar 0,99 menggunakan konfigurasi hyperparameter C = 100 dan gamma = 0,01. Keunggulan SVM terletak pada kemampuannya memodelkan hubungan non-linier dan interaksi kompleks antar fitur pada data dengan batas keputusan kelas yang tumpang tindih. Sebaliknya, RF menunjukkan kecenderungan over-prediction terhadap kelas minoritas ('Low') akibat mekanisme class_weight='balanced' dan keterbatasan pemisahan berbasis sumbu. Temuan ini menegaskan bahwa SVM dengan kernel Radial Basis Function (RBF) lebih sesuai untuk klasifikasi tingkat keparahan serangan siber, khususnya pada sistem deteksi insiden otomatis yang menuntut keseimbangan presisi dan recall serta pengambilan keputusan yang andal.*

ABSTRACT. *The escalating volume and sophistication of cyberattacks on network infrastructures processing massive daily traffic have overwhelmed security teams in prioritizing incident responses rapidly and accurately, a phenomenon known as alert fatigue. This study aims to analyze and compare the performance of the Support Vector Machine (SVM) and Random Forest (RF) algorithms for classifying cyberattack severity levels (Low, Medium, and High). The study uses the public Cyber Security Attacks dataset, consisting of 40,000 network traffic records reduced to 20,000 clean entries through preprocessing and feature engineering. The methodology includes data cleaning, selecting 10 significant features using SelectKBest, standardizing numerical features, and evaluating models across three data split scenarios (70:30, 80:20, and 90:10) using a stratified splitting approach. Experimental results show that SVM consistently outperforms RF across all scenarios, with the best performance in the 80:20 split, achieving 98.92% accuracy and a weighted average F1-Score of 0.99 using hyperparameter configurations of C = 100 and gamma = 0.01. The superiority of SVM lies in its ability to model non-linear relationships and complex feature interactions in data with overlapping class boundaries. In contrast, RF exhibits an over-prediction bias toward the minority class ('Low') due to the class_weight='balanced' mechanism and limitations of axis-based separation. These findings confirm that SVM with a Radial Basis Function (RBF) kernel is more suitable for cyberattack severity classification, particularly in automated incident detection systems requiring balanced precision and recall as well as reliable decision-making.*

*Penulis Korespondensi.

Email : *sri.winiarti@tif.uad.ac.id* (S. Winiarti)
Homepage : http://ejurnal.ung.ac.id/index.php/euler/index / E-ISSN : 2776-3706

1. Pendahuluan

Lanskap ancaman siber modern telah berevolusi menjadi sangat kompleks dengan dominasi serangan canggih seperti *Advanced Persistent Threats* (APT) dan teknik penyamaran (*evasion techniques*) yang memanfaatkan kerentanan *zero-day* [1–4]. Peningkatan volume dan kecanggihan serangan ini menyebabkan pendekatan pertahanan konvensional, khususnya Sistem Deteksi Intrusi (IDS) berbasis tanda tangan (*signature-based*), menjadi tidak lagi memadai karena ketidakmampuannya mendeteksi varian serangan baru atau polimorfik [5, 6]. Kegagalan deteksi ini tidak hanya meningkatkan risiko keamanan, tetapi juga membebani tim operasional dengan ribuan peringatan palsu atau tidak relevan (*alert fatigue*), yang menghambat proses triase insiden [7]. Oleh karena itu, kebutuhan akan mekanisme klasifikasi tingkat keparahan (*severity*) insiden secara otomatis yang akurat menjadi sangat krusial untuk memprioritaskan respons keamanan secara efektif [8].

Machine learning (ML) menawarkan solusi adaptif untuk mengatasi keterbatasan tersebut. Di antara berbagai algoritma, *Random Forest* (RF) dan *Support Vector Machine* (SVM) secara konsisten menunjukkan kinerja unggul dalam domain keamanan siber; RF dikenal karena ketahanannya terhadap *overfitting*, sedangkan SVM unggul dalam menangani data berdimensi tinggi [9–12]. Meskipun demikian, efektivitas kedua model ini sangat bergantung pada konfigurasi *hyperparameter* yang tepat. Tinjauan literatur menunjukkan adanya kesenjangan metodologis yang signifikan dalam penelitian sebelumnya. Studi yang ada cenderung terfragmentasi: sebagian berfokus pada optimasi mendalam namun terbatas pada satu jenis model saja [13–15], sedangkan studi yang membandingkan RF dan SVM sering kali mengabaikan proses optimasi konfigurasi yang cermat atau tidak menjadikannya sebagai fokus utama [15–17].

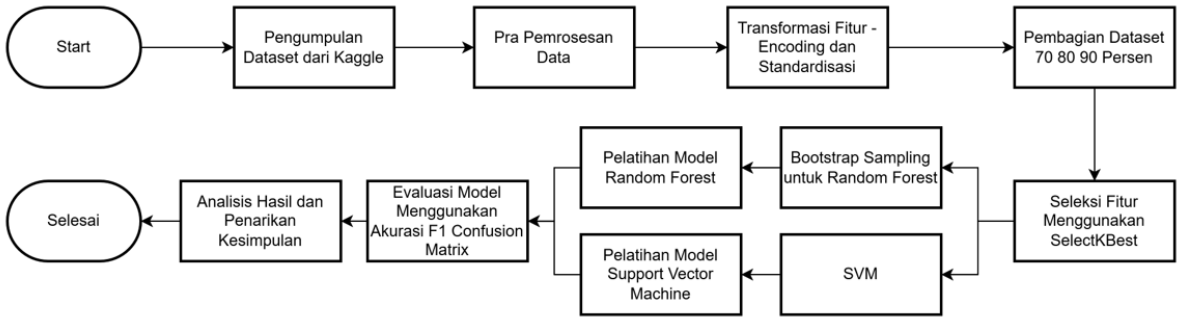
Kondisi ini menyisakan dua lapis kesenjangan yang saling memperkuat. Pertama, dari sisi metodologi: masih terbatasnya penelitian yang secara adil mempertandingkan RF dan SVM dalam kondisi performa puncak masing-masing (*optimized state*); studi komparatif yang ada umumnya tidak menerapkan strategi optimasi *hyperparameter* yang setara pada kedua sisi, sehingga kesimpulan mengenai superioritas satu model menjadi tidak valid [16–18]. Kedua, dari sisi karakteristik data: sebagian besar studi terdahulu mengevaluasi model pada dataset dengan label kelas biner (normal vs. attack) atau menggunakan dataset *benchmark* standar seperti *NSL-KDD* yang sudah seimbang [13, 14, 19]. Dataset *Cyber Security Attacks* yang digunakan dalam penelitian ini menghadirkan tantangan yang berbeda secara fundamental: tugas klasifikasi tingkat keparahan multi-kelas ordinal (Low, Medium, High) dengan distribusi kelas yang tidak seimbang secara alami (kelas Low hanya 16,12%), serta variabel target yang memerlukan rekayasa ulang melalui aturan heuristik sebelum dapat digunakan, kondisi yang lebih representatif terhadap skenario operasional nyata namun belum banyak dieksplorasi dalam literatur komparatif RF vs. SVM.

Penelitian ini bertujuan untuk mengisi kedua kesenjangan tersebut secara simultan. Kontribusi utamanya adalah: (1) penerapan strategi optimasi *hyperparameter* yang sistematis dan setara pada kedua model untuk mengekstraksi performa maksimal masing-masing, dan (2) penyediaan bukti empiris mengenai model mana yang lebih unggul dalam menangani kompleksitas klasifikasi keparahan serangan siber multi-kelas pada kondisi distribusi kelas yang tidak seimbang.

2. Metode

Penelitian ini menerapkan kerangka kerja eksperimental kuantitatif yang dirancang secara sistematis untuk mengevaluasi dan membandingkan kinerja model *machine learning*. Alur kerja penelitian dimulai dari pengumpulan data, pra-pemrosesan, pelatihan model, hingga

evaluasi kinerja, sebagaimana diilustrasikan pada Gambar 1.



Gambar 1. Diagram Alir Tahapan Penelitian

Penelitian ini membandingkan dua algoritma klasifikasi dengan mekanisme pembelajaran yang fundamental berbeda. *Random Forest* (RF) merupakan metode *ensemble* berbasis *bagging* yang membangun sejumlah pohon keputusan secara paralel, dikenal karena ketahanannya terhadap *overfitting* dan kemampuan mengukur pentingnya fitur [18–20]. *Support Vector Machine* (SVM) bekerja dengan menemukan *hyperplane* optimal yang memaksimalkan margin antar kelas, dan dengan kernel *Radial Basis Function* (RBF) mampu memodelkan hubungan non-linier dalam ruang berdimensi tinggi [21–25]. Pemilihan kedua algoritma ini bersifat strategis karena keduanya telah terbukti kompetitif dalam domain keamanan siber, namun belum pernah dibandingkan secara langsung dalam kondisi *hyperparameter* yang telah dioptimalkan untuk kasus klasifikasi keparahan multi-kelas [6, 13, 21, 22]. Berdasarkan Gambar 1, tahapan penelitian dijelaskan sebagai berikut.

2.1. Data dan Pra-Pemrosesan

Tahap awal dalam metodologi penelitian ini berfokus pada persiapan data, yang mencakup akuisisi dataset dan serangkaian prosedur pra-pemrosesan yang ketat untuk memastikan kualitas dan relevansi data sebelum tahap permodelan.

2.2. Akuisisi dan Karakteristik Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset publik berjudul "*Cyber Security Attacks*" yang disediakan oleh Incribo melalui platform *Kaggle* dan dapat diakses pada tautan berikut: https://www.kaggle.com/datasets/teamincribo/cyber-security-attacks/data. Dataset ini bersumber dari repositori *GitHub* Incribo dan mencakup simulasi lalu lintas jaringan komputer yang representatif untuk skenario keamanan siber.

Dalam kondisi awal, dataset terdiri dari 40.000 entri rekaman dengan 25 fitur variabel yang mencakup informasi lalu lintas jaringan secara komprehensif, di antaranya *Timestamp*, alamat IP sumber dan tujuan, protokol jaringan, panjang paket (*packet length*), indikator malware, skor anomali (*anomaly scores*), serta label tingkat keparahan (*Severity Level*) yang menjadi fokus utama dalam studi ini.

Variabel target *Severity Level* kemudian direkayasa ulang melalui proses *re-labelling* menjadi variabel baru '*Severity Level New*' dengan tiga kategori kelas ordinal: 'Low', 'Medium', dan 'High'. Rekayasa ini dilakukan berdasarkan serangkaian aturan heuristik yang menggabungkan tiga fitur utama, yaitu *Anomaly Scores*, *Action Taken*, dan *Packet Length*. Setelah melalui tahap pembersihan data, termasuk penghapusan delapan kolom non-prediktif seperti *Timestamp* dan IP Address serta imputasi nilai yang hilang, jumlah entri yang valid dan siap digunakan dalam proses pemodelan adalah 20.000 entri. Proses rekayasa variabel target ini menghasilkan dis-

tribusi kelas yang tidak seimbang (*imbalanced*), sebagaimana diilustrasikan pada Gambar 2, yang menjadi pertimbangan utama dalam pemilihan strategi evaluasi pada penelitian ini.

2.3. Prosedur Pra-pemrosesan dan Rekayasa Fitur

Tahap pra-pemrosesan dirancang untuk mengubah data mentah menjadi representasi fitur yang optimal bagi kedua algoritma klasifikasi. Prosedur ini dilakukan secara berurutan melalui empat langkah utama.

1. Pembersihan Data (*Data Cleaning*) dilakukan sebagai langkah pertama dengan menghapus delapan kolom non-prediktif yang tidak berkontribusi pada proses klasifikasi, yaitu *Timestamp*, *Source IP Address*, *Destination IP Address*, *Source Port*, *Destination Port*, *Payload Data*, *Firewall Logs*, dan *IDS/IPS Alerts*. Selain itu, nilai yang hilang (*missing values*) pada kolom numerik diimputasi menggunakan nilai median kelas yang bersangkutan untuk menghindari distorsi distribusi data. Proses ini mereduksi dataset dari 40.000 entri awal menjadi 20.000 entri yang valid dan bersih.
2. Rekayasa Variabel Target (*Feature Engineering*) dilakukan melalui proses *re-labelling* variabel *Severity Level* menjadi variabel baru '*Severity Level New*' dengan tiga kategori kelas ordinal: 'Low', 'Medium', dan 'High'. Transformasi ini menggunakan aturan heuristik yang menggabungkan nilai *Anomaly Scores*, *Action Taken*, dan *Packet Length* sebagai dasar penentuan kelas keparahan.
3. Seleksi Fitur (*Feature Selection*) dilakukan menggunakan metode *SelectKBest* dengan uji statistik *Analysis of Variance* (ANOVA *F-test*), yang mengukur hubungan linier antara setiap fitur numerik dengan variabel target. Dari 21 fitur yang tersedia setelah pembersihan data, dipilih 10 fitur dengan skor statistik tertinggi. Reduksi dimensi ini bertujuan menghilangkan fitur yang tidak relevan secara statistik dan mengurangi risiko multikolinearitas, yang dapat mendistorsi performa khususnya pada SVM yang sensitif terhadap skala dan dimensi ruang fitur. Persamaan uji ANOVA *F-test* dinyatakan dalam pers. (1).
$$F = \frac{\text{variasi antar kelompok}}{\text{variasi dalam kelompok}} = \frac{MS_{\text{between}}}{MS_{\text{within}}}, \tag{1}$$
dimana nilai F yang lebih tinggi mengindikasikan fitur yang lebih diskriminatif terhadap kelas target [26].
4. Standarisasi Fitur (*Feature Standardization*) dilakukan menggunakan *StandardScaler* dari library *Scikit-Learn*, yang mentransformasi setiap fitur numerik agar memiliki rata-rata (*mean*) = 0 dan standar deviasi = 1, sebagaimana dinyatakan dalam pers. (2).
$$z = \frac{x - \mu}{\sigma}, \tag{2}$$
dimana x adalah nilai fitur asli, μ adalah rata-rata fitur, dan σ adalah standar deviasinya. Standarisasi ini krusial khususnya bagi SVM, karena algoritma tersebut sangat sensitif terhadap perbedaan skala antar fitur dalam pembentukan *hyperplane* optimal.

2.4. Konfigurasi Operasional dan Optimasi Model

Untuk menjawab tantangan klasifikasi pada data serangan siber yang memiliki karakteristik non-linier, penelitian ini menerapkan strategi konfigurasi *hyperparameter* yang berbeda untuk kedua algoritma. Rincian konfigurasi operasional disajikan pada Tabel 1.

Seperti dirangkum pada Tabel 1, strategi konfigurasi yang diterapkan dalam penelitian ini bersifat eksploratif-bertingkat. Untuk *Random Forest*, tiga konfigurasi diuji mulai dari model paling sederhana (RF-Original dengan max_depth = 3) hingga model paling kompleks

Tabel 1. Konfigurasi *Hyperparameter* dan Strategi Optimasi

Algoritma	Skenario	Nama Konfigurasi	Parameter Utama
Random Forest	1	RF-Original	n_estimators = 100, max_depth = 3, class_weight = 'balanced'
	2	RF-Controlled	n_estimators = 100, max_depth = 12, class_weight = 'balanced'
	3	RF-Complex	n_estimators = 200, max_depth = 20, class_weight = 'balanced'
Support Vector Machine	1	SVM-Baseline	C = 1, gamma = 'scale', kernel = 'rbf'
	2	SVM-Controlled	C = 100, gamma = 0,01, kernel = 'rbf'
	3	SVM-High C	C = 1000, gamma = 0,001, kernel = 'rbf'

(RF-Complex dengan max_depth = 20 dan n_estimators = 200). Pembatasan kedalaman pada RF-Original dilakukan secara sengaja (*controlled depth*) untuk mensimulasikan model ringkas sekaligus menguji hipotesis bahwa struktur pohon yang sederhana tidak cukup memadai dalam menangkap kompleksitas pola keparahan serangan siber dibandingkan mekanisme *hyperplane* pada SVM.

Untuk *Support Vector Machine*, tiga konfigurasi kernel RBF diuji dengan rentang nilai C yang bervariasi dari konfigurasi default *Scikit-Learn* (SVM-Baseline, C = 1) hingga penalti yang sangat tinggi (SVM-High C, C = 1000). Konfigurasi SVM-Controlled dengan C = 100 dan gamma = 0,01 terpilih sebagai konfigurasi optimal karena menghasilkan keseimbangan terbaik antara fleksibilitas batas keputusan dan kemampuan generalisasi pada data uji.

Perlu dicatat bahwa ketiga skenario konfigurasi pada Tabel 1 merupakan eksperimen perbandingan untuk mengidentifikasi konfigurasi terbaik secara konseptual. Adapun konfigurasi yang diterapkan pada masing-masing skenario pembagian data (70:30, 80:20, dan 90:10) ditentukan secara independen melalui proses *Grid Search Cross-Validation*, sehingga parameter optimal yang terpilih dapat berbeda antar skenario pembagian sebagaimana tercermin pada hasil eksperimen pada Tabel 2.

Kriteria pemilihan model terbaik (*Best Model Selection*) secara keseluruhan didasarkan pada nilai *F1-Score* (*Weighted Average*) tertinggi yang dicapai pada skenario evaluasi 80:20, mengingat metrik ini memberikan penilaian yang paling adil terhadap dataset dengan distribusi kelas yang tidak seimbang.

2.5. Strategi Evaluasi dan Validasi Model

Untuk menjamin validitas hasil eksperimen pada dataset serangan siber yang memiliki karakteristik distribusi kelas tidak seimbang (*imbalanced*), penelitian ini menerapkan strategi evaluasi yang ketat meliputi teknik pembagian data dan pemilihan metrik prioritas.

2.5.1. Teknik Pembagian Data (*Stratified Data Splitting*)

Alih-alih menggunakan pembagian acak sederhana (*simple random sampling*), penelitian ini menerapkan teknik *Stratified Shuffle Split*. Teknik ini dipilih untuk memastikan bahwa proporsi setiap kelas target (Low, Medium, High) pada data latih (*training set*) dan data uji (*testing set*) tetap konsisten dengan distribusi pada dataset asli [27].

Hal ini krusial untuk mencegah fenomena *sampling bias*, di mana kelas minoritas (khususnya serangan tingkat 'Low' dengan proporsi 16,1%) mungkin kurang terwakili dalam data uji jika menggunakan metode acak, yang dapat menyebabkan evaluasi performa menjadi tidak valid. Tiga skenario rasio pembagian diuji, yaitu 70:30, 80:20, dan 90:10, untuk mengukur konsistensi (*robustness*) model terhadap variasi ketersediaan volume data latih [15].

2.5.2. Pemilihan Metrik Evaluasi Utama

Meskipun Akurasi sering digunakan sebagai metrik standar, penggunaannya pada dataset tidak seimbang dapat memberikan gambaran yang menyesatkan (*accuracy paradox*). Sebuah model dapat mencapai akurasi tinggi hanya dengan memprediksi kelas mayoritas (High dan Medium) secara benar namun gagal mengenali pola kelas minoritas (Low) secara akurat, sehingga memberikan gambaran kinerja yang menyesatkan. Oleh karena itu, penelitian ini menetapkan *F1-Score* (*Weighted Average*) sebagai metrik evaluasi primer dengan rumus sebagai berikut:

$$F1_{\text{weighted}} = \sum_{i=1}^{n} (w_i \times F1_i), \tag{3}$$

dimana w_i adalah proporsi jumlah sampel untuk kelas ke-i. Penggunaan rata-rata tertimbang (*weighted*) memastikan bahwa kinerja model pada kelas minoritas tetap berkontribusi signifikan terhadap skor akhir. Selain itu, dasar perhitungan metrik ini tetap mengacu pada komponen *Confusion Matrix* yang terdiri dari *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) untuk memetakan kesalahan klasifikasi spesifik antar tingkat keparahan serangan [24–26].

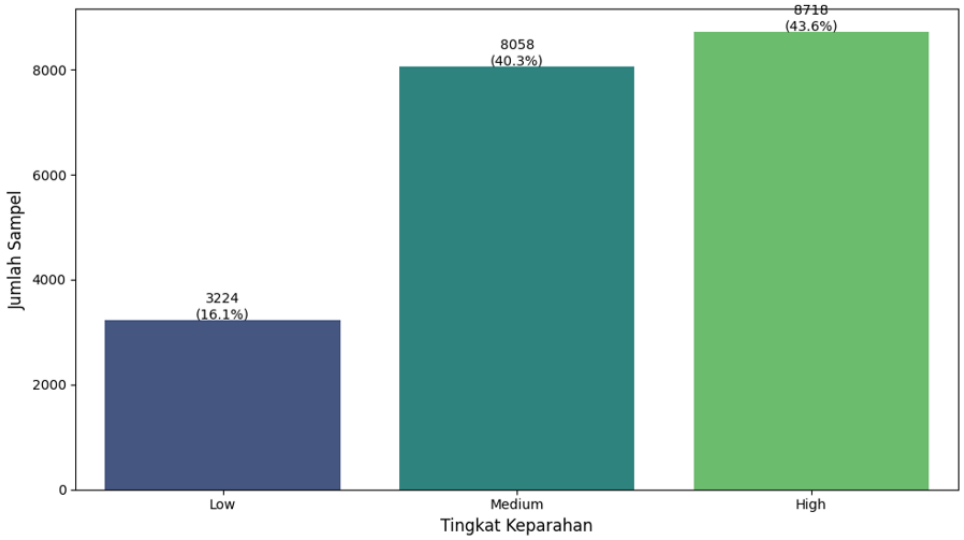
3. Hasil dan Pembahasan

Bagian ini menyajikan temuan empiris yang diperoleh dari rangkaian eksperimen komputasional, dimulai dari hasil transformasi data hingga evaluasi kinerja model klasifikasi. Hasil yang dipaparkan tidak hanya bersifat deskriptif, melainkan disertai dengan analisis mendalam untuk menginterpretasikan fenomena yang terjadi pada data dan perilaku algoritma. Diskusi difokuskan pada pembuktian hipotesis mengenai efektivitas optimasi *hyperparameter* dalam meningkatkan akurasi deteksi serangan siber, serta perbandingan kritis antara arsitektur *Random Forest* (RF) dan *Support Vector Machine* (SVM) dalam menangani kompleksitas data serangan multi-kelas.

3.1. Karakteristik Data Hasil Pra-pemrosesan

Tahapan pra-pemrosesan yang telah dijelaskan pada Bagian 2 menghasilkan dataset akhir sebanyak 20.000 entri bersih dengan 10 fitur terpilih dan tiga kelas target ordinal: ‘Low’, ‘Medium’, dan ‘High’. Temuan penting dari tahap ini adalah adanya distribusi kelas yang tidak seimbang (*imbalanced*) pada variabel target, sebagaimana ditunjukkan pada Gambar 2. Ketidakseimbangan ini menjadi landasan utama pemilihan *F1-Score* (*weighted average*) sebagai metrik evaluasi primer, karena metrik akurasi tunggal dapat memberikan gambaran yang menyesatkan pada kondisi distribusi kelas yang tidak merata.

Gambar 2 memperlihatkan distribusi kelas pada variabel target ‘*Severity Level New*’ hasil rekayasa fitur dari 20.000 entri data yang telah melalui proses pembersihan. Distribusi menunjukkan ketimpangan yang nyata antar kelas: kelas ‘High’ mendominasi dengan 8.718 sampel (43,60%), diikuti kelas ‘Medium’ sebesar 8.058 sampel (40,30%), sementara kelas ‘Low’ menjadi minoritas dengan hanya 3.224 sampel (16,10%). Kondisi ketidakseimbangan ini khususnya pada kelas ‘Low’ yang hanya mencakup seperenam dari total data menjadi tantangan tersendiri bagi model klasifikasi karena berisiko menghasilkan *under-representation* pada kelas minoritas dalam proses pelatihan model. Oleh karena itu, sebagaimana telah ditetapkan pada Bagian 2.5, *F1-Score* (*weighted average*) dipilih sebagai metrik evaluasi primer untuk memastikan kontribusi kelas minoritas tetap diperhitungkan secara proporsional dalam penilaian kinerja model.



Gambar 2. Distribusi Kelas Keparahan Serangan pada Dataset Final

3.2. Hasil Eksperimen Model

Evaluasi kinerja model dilakukan secara sistematis melalui tiga skenario pembagian data latih dan uji (70:30, 80:20, dan 90:10) menggunakan teknik *stratified splitting* untuk menjaga proporsi kelas. Hasil eksperimen kuantitatif yang mencakup metrik *F1-Score* (*weighted average*) dan Akurasi dari kedua model dirangkum secara komprehensif pada Tabel 2.

Tabel 2. Perbandingan Kinerja SVM dan RF pada Data Uji

Algoritma	Skenario	Akurasi	Presisi	Recall	F1-Score
Random Forest	70:30	91,03%	0,94	0,91	0,91
	80:20	91,07%	0,94	0,91	0,92
	90:10	91,15%	0,94	0,91	0,92
Support Vector Machine	70:30	98,75%	0,99	0,99	0,99
	80:20	98,92%	0,99	0,99	0,99
	90:10	97,10%	0,97	0,97	0,97

Berdasarkan data pada Tabel 2, terlihat pola kinerja yang sangat konsisten di mana model *Support Vector Machine* (SVM) menunjukkan superioritas yang signifikan dibandingkan *Random Forest* (RF) di seluruh skenario pengujian. Pada skenario 80:20, SVM mencapai performa puncaknya dengan *F1-Score* sebesar 0,99 dan Akurasi 98,92%, sebuah angka yang mengindikasikan kemampuan generalisasi yang sangat baik pada data yang belum pernah dilihat sebelumnya. Sebaliknya, meskipun RF menunjukkan stabilitas dengan *F1-Score* berkisar antara 0,91 hingga 0,92, model ini gagal melampaui ambang batas performa yang ditetapkan oleh SVM.

Lebih lanjut, variasi kinerja SVM antar skenario menunjukkan pola yang menarik, di mana kinerjanya tidak mengikuti tren peningkatan yang konsisten seiring bertambahnya proporsi data latih. SVM mencapai performa terbaiknya pada skenario 80:20 dengan akurasi 98,92% dan *F1-Score* 0,99, namun mengalami penurunan yang cukup signifikan pada skenario 90:10 dengan akurasi 97,10% dan *F1-Score* 0,97, lebih rendah dibandingkan skenario 70:30 yang mencapai akurasi 98,75%. Penurunan ini bukan merupakan anomali acak, melainkan konsekuensi langsung dari perbedaan konfigurasi *hyperparameter* yang diperoleh melalui proses optimasi pada masing-masing skenario.

Pada skenario 70:30 dan 80:20, proses optimasi menghasilkan konfigurasi dengan nilai C yang relatif besar ($C = 10$ dan $C = 100$), yang memungkinkan model membentuk batas kepu-

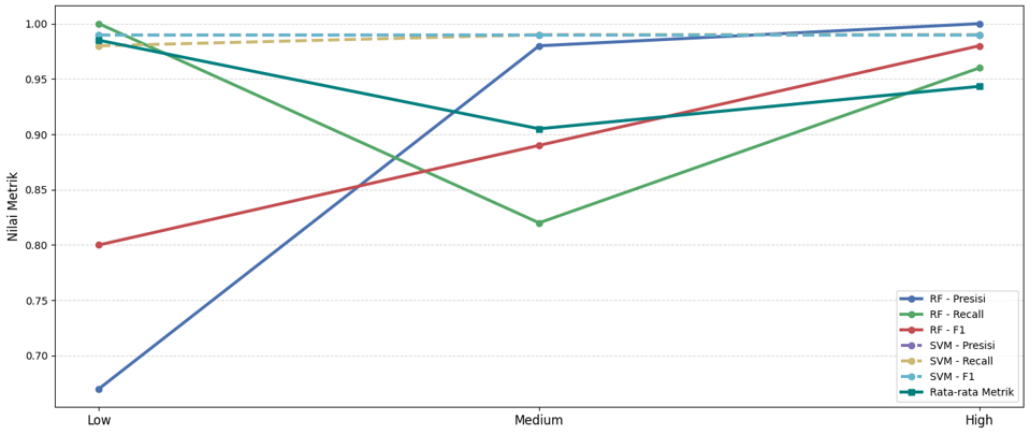
tusan yang lebih fleksibel dan adaptif terhadap kompleksitas data serangan siber. Sebaliknya, pada skenario 90:10, proses optimasi justru menghasilkan nilai C = 0,1, sebuah konfigurasi regularisasi yang sangat ketat yang secara signifikan membatasi kapasitas model dalam memodelkan hubungan non-linier antar fitur. Temuan ini menggarisbawahi bahwa pada SVM, volume data latih yang lebih besar tidak secara otomatis menjamin kinerja yang lebih tinggi apabila tidak disertai dengan konfigurasi *hyperparameter* yang sesuai. Oleh karena itu, konfigurasi 80:20 dengan C = 100 dan gamma = 0,01 ditetapkan sebagai konfigurasi optimal dalam penelitian ini.

3.3. Analisis Kualitatif dan Interpretasi Model

Untuk memahami dinamika di balik perbedaan kinerja kuantitatif yang signifikan, analisis mendalam dilakukan dengan membedah performa per kelas. Rincian nilai metrik untuk setiap kelas pada skenario terbaik (80:20) disajikan dalam Tabel 3, sementara perbandingan visual pola kinerjanya diilustrasikan pada Gambar 3.

Tabel 3. Perbandingan Kinerja Per Kelas pada Skenario 80:20

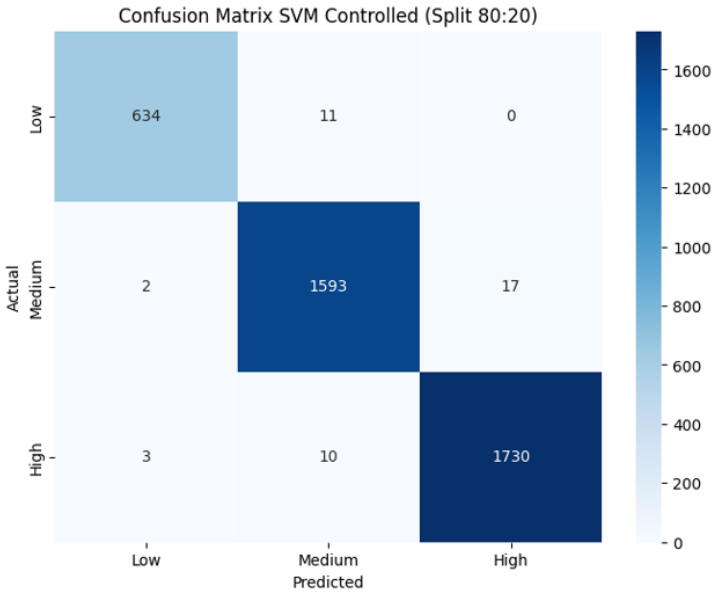
Model	Kelas	Presisi	Recall	F1-Score
Random Forest	Low	0.67	1.00	0.80
	Medium	0.98	0.82	0.89
	High	1.00	0.96	0.98
Support Vector Machine	Low	0.99	0.98	0.99
	Medium	0.99	0.99	0.99
	High	0.99	0.99	0.99



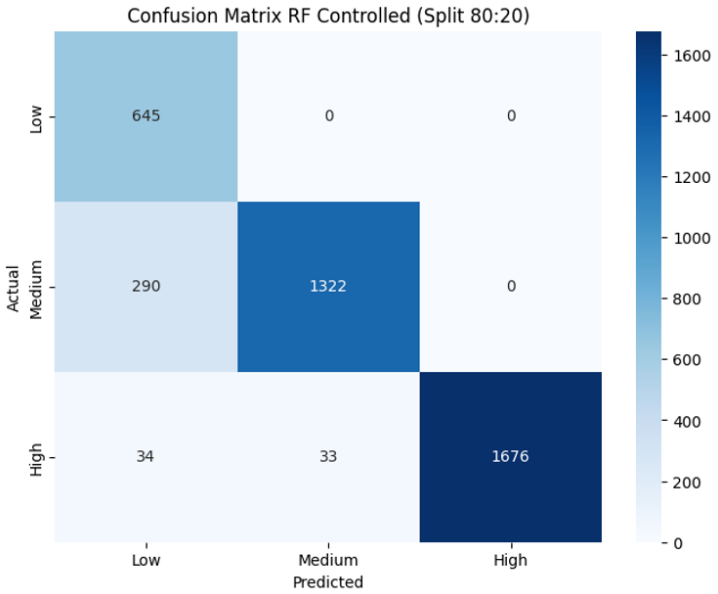
Gambar 3. Visualisasi Perbandingan Presisi, *Recall*, dan *F1-Score* per Kelas

Berdasarkan Tabel 2 dan Gambar 3, terlihat anomali menarik pada model RF. Model ini mencapai nilai *Recall* sempurna (1.00) untuk kelas 'Low', yang berarti ia berhasil mendeteksi seluruh serangan bertipe rendah tanpa ada yang terlewat. Namun, pencapaian ini diikuti oleh nilai Presisi yang rendah (0.67). Disparitas visual yang tajam antara batang *Recall* (hijau) dan Presisi (biru) pada kelas 'Low' pada Gambar 3 menegaskan bahwa model RF cenderung melakukan *over-prediction*, di mana banyak sampel dari kelas lain secara keliru diklasifikasikan sebagai 'Low'.

Sebaliknya, model SVM menunjukkan keseimbangan yang luar biasa. Baik pada tabel maupun grafik, nilai Presisi, *Recall*, dan *F1-Score* terlihat seragam dan mendekati angka sempurna di ketiga kelas. Hal ini mengindikasikan bahwa SVM berhasil menangani ketidakseimbangan distribusi kelas secara proporsional, tanpa kecenderungan *over-prediction* pada kelas tertentu.



Gambar 4. *Confusion Matrix* Model *Support Vector Machine* (SVM) pada Skenario 80:20



Gambar 5. *Confusion Matrix* Model *Random Forest* (RF) pada Skenario 80:20

Pola ini dikonfirmasi secara visual oleh *confusion matrix* pada Gambar 4 dan Gambar 5. Tingginya angka *False Positives* pada kolom kelas 'Low' pada Gambar 5 membuktikan bahwa RF cenderung *over-predict* terhadap kelas minoritas sebagai konsekuensi langsung dari konfigurasi *class_weight='balanced'* yang dikombinasikan dengan pembatasan kedalaman pohon. Sebaliknya, diagonal yang bersih pada Gambar 4 menunjukkan bahwa SVM berhasil memisahkan ketiga kelas secara presisi dan merata di ruang berdimensi tinggi.

Perbandingan dengan studi terdahulu memperkuat validitas temuan ini dan menggarisbawahi kontribusi spesifik dari pendekatan penelitian ini. Elashmawi, dkk. [15] yang mengevaluasi RF, SVM, dan *Decision Tree* pada dataset *NSL-KDD* menggunakan skenario tiga kelas melaporkan bahwa RF mencapai akurasi tertinggi sebesar 85,42% pada data uji, sekaligus mengungguli SVM dalam kondisi tersebut. Namun, *NSL-KDD* merupakan dataset *benchmark* yang relatif seimbang dan berlabel biner antara lalu lintas normal dan serangan. Berbeda se-

cara fundamental dengan tantangan klasifikasi keparahan multi-kelas ordinal yang dihadapi dalam penelitian ini. Temuan Elashmawi, dkk. justru memperkuat argumen bahwa keunggulan RF bersifat konteks-spesifik: pada dataset yang lebih sederhana dan seimbang, RF mampu mengungguli SVM, namun pada dataset dengan distribusi kelas yang tidak seimbang dan batas keputusan non-linier yang kompleks seperti pada penelitian ini, SVM dengan kernel *Radial Basis Function* (RBF) memiliki keunggulan yang jauh lebih signifikan.

Senada dengan hal tersebut, Pandey, dkk. [13] yang mengoptimalkan RF menggunakan metode *Tabu Search* pada dataset *WSN-DS* dan *CICIDS 2017* mencatat bahwa SVM tanpa optimasi hanya mampu mencapai akurasi 90,8% dalam tugas deteksi intrusi, sementara RF yang dioptimalkan mampu melampaui angka tersebut secara signifikan. Temuan ini selaras dengan hasil penelitian ini: ketika kedua model sama-sama melewati proses optimasi *hyperparameter* yang setara dan sistematis melalui *Grid Search Cross-Validation*, SVM justru menunjukkan keunggulan yang lebih besar dibandingkan RF, yang mengindikasikan bahwa SVM memiliki kapasitas ekstraksi yang lebih tinggi ketika potensinya dimaksimalkan secara optimal pada data serangan siber dengan karakteristik non-linier. Nanda, dkk. [28] yang membandingkan RF dan SVM pada dataset intrusi jaringan tanpa proses optimasi terstruktur melaporkan selisih kinerja yang lebih sempit antara kedua model, semakin menegaskan bahwa kesenjangan performa besar yang teridentifikasi dalam penelitian ini merupakan cerminan langsung dari efek optimasi *hyperparameter* yang setara pada kedua sisi.

Perlu dicatat bahwa performa SVM yang sangat tinggi (akurasi 98,92% dan *F1-Score* 0,99) tidak mengindikasikan *overfitting*, mengingat evaluasi seluruhnya dilakukan pada data uji (*testing set*) yang sepenuhnya terpisah dari data latih (*training set*) dan tidak pernah digunakan dalam proses pelatihan maupun optimasi *hyperparameter*. Konsistensi kinerja SVM pada ketiga skenario pembagian data meskipun dengan nilai C yang berbeda-beda hasil *Grid Search* semakin mengonfirmasi bahwa model telah berhasil melakukan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya, bukan sekadar menghafal pola data latih.

Temuan ini menunjukkan bahwa SVM dengan kernel RBF berpotensi menjadi kandidat yang lebih kuat untuk modul klasifikasi tingkat keparahan serangan dalam sistem deteksi insiden otomatis, khususnya pada lingkungan operasional yang menghadapi tantangan *alert fatigue* dan menuntut keseimbangan presisi tinggi di seluruh kelas keparahan.

Keunggulan SVM dalam penelitian ini merupakan hasil dari interaksi dua faktor yang saling melengkapi: pertama, kesesuaian intrinsik antara mekanisme kernel *Radial Basis Function* (RBF) dan karakteristik data serangan siber yang bersifat non-linier dan multi-kelas; dan kedua, efektivitas konfigurasi *hyperparameter* yang optimal, yakni C = 100 dan gamma = 0,01, dalam mengekstraksi kapasitas diskriminatif maksimal dari kernel tersebut. Tanpa konfigurasi yang tepat, potensi kernel RBF tidak akan terwujud sepenuhnya; sebaliknya, tanpa kesesuaian karakteristik data, optimasi *hyperparameter* sekalipun tidak akan menghasilkan performa yang sedemikian superior. Secara spesifik, data serangan siber dalam penelitian ini menunjukkan pola non-linier dan tumpang tindih antar kelas keparahan; kelas 'High' tidak membentuk satu klaster tunggal dalam ruang fitur, melainkan tersebar dalam beberapa pola perilaku yang berbeda, seperti serangan berdurasi panjang dengan intensitas lalu lintas sedang maupun serangan berdurasi singkat dengan lonjakan lalu lintas yang sangat tinggi. Kondisi inilah yang menyebabkan batas keputusan antar kelas tidak dapat dipisahkan secara linier, sehingga kernel RBF yang memproyeksikan data ke ruang berdimensi lebih tinggi menjadi mekanisme yang paling sesuai untuk menangkap kompleksitas tersebut.

Sebaliknya, meskipun *Random Forest* dikenal sebagai model yang kuat dan tahan terhadap *overfitting*, hasil penelitian menunjukkan bahwa RF mengalami bias prediksi yang signifikan

terhadap kelas minoritas ('Low'), yang hanya mencakup 16,1% dari total dataset. Penggunaan parameter *class_weight='balanced'* menyebabkan model memberikan bobot pelatihan yang lebih besar pada kelas ini, menghasilkan nilai *Recall* sempurna (1,00) namun diikuti oleh *Precision* yang rendah (0,67), mengindikasikan kecenderungan *over-prediction* di mana banyak sampel dari kelas Medium dan High salah diklasifikasikan sebagai Low. Fenomena ini diperparah oleh mekanisme pemisahan berbasis sumbu (*axis-aligned splits*) pada pohon keputusan yang kurang fleksibel dalam menangkap batas keputusan non-linier antar kelas keparahan, serta pembatasan kedalaman pohon yang membatasi kapasitas RF dalam membedakan pola yang halus dan saling tumpang tindih.

Analisis *confusion matrix* semakin memperkuat temuan ini. Pada model RF, kesalahan klasifikasi terutama terjadi ketika sampel dari kelas Medium dan High diprediksi sebagai Low, yang dalam konteks operasional keamanan siber berpotensi berbahaya karena dapat menurunkan prioritas penanganan insiden yang sebenarnya kritis. Sebaliknya, SVM menunjukkan pola kesalahan yang minimal dan distribusi prediksi yang seimbang di seluruh kelas, menandakan kemampuan generalisasi yang lebih baik terhadap data yang belum pernah dilihat sebelumnya.

Temuan ini memiliki implikasi praktis yang signifikan bagi pengembangan sistem deteksi insiden otomatis. Dalam lingkungan operasional yang menghadapi masalah *alert fatigue*, presisi yang tinggi dan keseimbangan prediksi antar kelas menjadi lebih penting dibandingkan sekadar *recall* tinggi pada satu kelas tertentu. Oleh karena itu, hasil penelitian ini mengindikasikan bahwa SVM dengan kernel RBF lebih sesuai untuk modul klasifikasi tingkat keparahan serangan, khususnya pada sistem yang menuntut akurasi tinggi dalam pengambilan keputusan dan alokasi sumber daya keamanan.

Secara keseluruhan, pembahasan ini menegaskan bahwa perbedaan kinerja antara SVM dan RF bukanlah fenomena kebetulan, melainkan konsekuensi langsung dari interaksi antara struktur data serangan siber dan mekanisme pembelajaran masing-masing algoritma. Dengan demikian, pemilihan model klasifikasi untuk keamanan siber sebaiknya tidak hanya didasarkan pada popularitas algoritma, tetapi juga pada pemahaman mendalam terhadap karakteristik data dan tujuan operasional sistem yang dibangun.

4. Kesimpulan

Support Vector Machine (SVM) dengan kernel *Radial Basis Function* (RBF) terbukti secara konsisten mengungguli *Random Forest* (RF) dalam seluruh skenario klasifikasi tingkat keparahan serangan siber multi-kelas yang diuji. Performa optimal dicapai pada skenario pembagian data 80:20 dengan konfigurasi *hyperparameter* C = 100 dan gamma = 0,01, menghasilkan akurasi sebesar 98,92% dan *F1-Score* tertimbang sebesar 0,99. Adapun penurunan kinerja SVM pada skenario 90:10 (akurasi 97,10%) bukan merupakan keterbatasan inheren algoritma, melainkan konsekuensi dari konfigurasi regularisasi yang lebih ketat (C = 0,10) yang dihasilkan oleh proses *Grid Search* pada proporsi data latih yang berbeda.

Keunggulan SVM dapat dijelaskan melalui dua faktor yang saling melengkapi: pertama, kesesuaian intrinsik antara kernel RBF dan karakteristik data serangan siber yang bersifat non-linier dengan batas keputusan antar kelas yang saling tumpang tindih; dan kedua, efektivitas strategi optimasi *hyperparameter* yang sistematis dalam mengekstraksi kapasitas diskriminatif maksimal dari model. Di sisi lain, RF menunjukkan keterbatasan dalam menangani distribusi kelas yang tidak seimbang khususnya *over-prediction* terhadap kelas minoritas ('Low', 16,10%) akibat mekanisme *class_weight='balanced'* yang berdampak pada penurunan presisi secara keseluruhan.

Temuan ini memiliki implikasi praktis yang jelas: pemilihan model klasifikasi keparahan serangan siber sebaiknya tidak hanya mempertimbangkan popularitas algoritma, tetapi juga kesesuaian struktural antara mekanisme model dan karakteristik data. Dalam konteks pengembangan sistem deteksi insiden otomatis yang menghadapi tantangan *alert fatigue*, SVM dengan kernel RBF dan konfigurasi yang dioptimalkan menunjukkan potensi sebagai kandidat yang lebih kuat untuk modul klasifikasi keparahan, dengan kemampuan generalisasi yang terbukti baik pada data uji yang sepenuhnya independen.

Sebagai arah pengembangan ke depan, penelitian selanjutnya disarankan untuk mengeksplorasi penerapan teknik *Explainable AI* (XAI) guna meningkatkan interpretabilitas model bagi praktisi keamanan, serta membandingkan pendekatan ini dengan arsitektur *deep learning* modern pada dataset yang lebih besar dan bersifat *real-time*.

Kontribusi Penulis. **Reyhanssan Islamey**: Validasi, analisis formal, investigasi, penulisan–persiapan draf asli. **Sri Winiarti**: Konseptualisasi, metodologi, perangkat lunak, penulisan–peninjauan dan penyuntingan. **Imam Riadi**: Penulisan–persiapan draf asli, sumber daya. Semua penulis mendiskusikan hasil dan berkontribusi pada manuskrip akhir.

Ucapan Terima Kasih. Kami ingin menyampaikan rasa terima kasih kami kepada semua pihak yang telah mendukung pelaksanaan penelitian kolaboratif internasional dengan skema hibah pendanaan bersama antara Universitas Ahmad Dahlan dan UMPSA pada tahun 2025, sesuai dengan kontrak No. 003/IRMG/LPPM-UAD/IX/2024. Kami juga ingin berterima kasih kepada Institut Penelitian dan Pengabdian Masyarakat Universitas Dahlan atas dukungan dan bimbingan mereka selama penelitian. Penulis juga berterima kasih kepada Evinda Apriliani atas diskusi produktif dan dukungan moral yang diberikan dalam penyusunan naskah ini.

Pembiayaan. Penelitian ini didanai oleh Universitas Ahmad Dahlan dan University Malaysia Pahang Sultan Abdullah dengan skema matching grant tahun 2025.

Konflik Kepentingan. Para penulis menyatakan tidak ada konflik kepentingan yang terkait dengan artikel ini.

Ketersediaan Data. Dataset yang digunakan dalam penelitian ini tersedia secara publik dan dapat diakses melalui repositori *Kaggle* (https://www.kaggle.com/datasets/teamincribo/cyber-security-attacks/data).

Referensi

[1] B. O. Zhang, Y. Gao, B. Kuang, C. Yu, A. Fu, and W. Susilo, "A survey on advanced persistent threat detection: A unified framework, challenges, and countermeasures," *ACM*, vol. 57, no. 3, 2026, doi: 10.1145/3700749.

[2] V. Sharma, "Advanced persistent threat (APT) detection using SIEM: A review of techniques and tools," *Engineering and Technology Journal*, vol. 10, no. 7, pp. 5738–5746, 2025, doi: 10.47191/etj/v10i07.21.

[3] A. Awaludin, W. Sulistyadi, and A. F. Chandra, "Analysis of attacks and cybersecurity in the health sector during the COVID-19 pandemic: A scoping review," *Journal of Social Science*, vol. 4, no. 1, pp. 62–70, Jan. 2023, doi: 10.46799/jss.v4i1.512.

[4] M. Bhukya *et al.*, "IoT network attack severity," in *E3S Web of Conferences*, 2023.

[5] A. S. Alqahtani, O. A. Altammami, and M. A. Haq, "A comprehensive analysis of network security attack classification using machine learning algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 4, pp. 1269–1280, 2024.

[6] H. Z. Rui, T. Y. Chien, L. X. Ee, C. W. San, and L. T. Yi, "Comparison of the use of support vector machine (SVM) and random forest (RF) algorithms for DDoS attack detection," *International Journal of Research and Innovation in Social Science*, vol. 9, pp. 1126–1138, 2025.

[7] M. Lopez-Ledezma, "Cyber security data science: Machine learning methods and their performance on imbalanced datasets," in *Digital Management and Artificial Intelligence*, 2024, pp. 1–13.
[8] R. Mai and M. Wu, "Using information technology to quantitatively evaluate and prevent cybersecurity threats in a hierarchical manner," *International Journal for Applied Information Management*, vol. 3, no. 1, pp. 1–10, 2023.
[9] R. Buchta, A. Data, and S. Hannover, "Advanced persistent threat attack detection systems: A review of approaches, challenges, and trends," *Digital Threats: Research and Practice*, vol. 5, no. 4, 2024, doi: 10.1145/3696014.
[10] D. Revaldo, "Implementation of random forest classification and support vector machine algorithms for phishing link detection," *Journal of Informatics, Information System, Software Engineering and Applications*, vol. 8106, pp. 127–137, 2024.
[11] V. Malik, A. Khanna, N. Sharma, and S. Nalluri, "Advanced persistent threats (APTs): Detection techniques and mitigation strategies," *International Journal of Global Innovations and Solutions*, 2024.
[12] F. Genuario *et al.*, "Machine learning-based methodologies for cyber-attacks and network traffic monitoring: A review and insights," *Information*, vol. 15, no. 11, 2024, doi: 10.3390/info15110741.
[13] V. K. Pandey *et al.*, "Enhancing intrusion detection in wireless sensor networks using a tabu search based optimized random forest," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-03498-3.
[14] Y. Chang and Y. Lin, "Support vector machines with hyperparameter optimization frameworks for classifying mobile phone prices in multi-class," *Electronics*, 2025.
[15] W. H. Elashmawi, A. Sheta, and A. Al-Qerem, "Intelligent intrusion detection system using RF, SVM, and DT: A comparison-based KDD dataset," *Journal of Computer Science*, vol. 21, no. 8, pp. 1749–1759, 2025, doi: 10.3844/jcssp.2025.1749.1759.
[16] B. Madhu *et al.*, "IoT network attack severity classification," in *E3S Web of Conferences*, 2023, doi: 10.1051/e3sconf/202343001152.
[17] W. Chen *et al.*, "A survey on imbalanced learning: Latest research, applications and future directions," *Artificial Intelligence Review*, 2024, doi: 10.1007/s10462-024-10759-6.
[18] I. H. Sarker *et al.*, "Cybersecurity data science: An overview from machine learning perspective," *Journal of Big Data*, 2020, doi: 10.1186/s40537-020-00318-5.
[19] S. T. Hamidou and A. Mehdi, "Enhancing IDS performance through a comparative analysis of random forest, XGBoost, and deep neural networks," *Machine Learning with Applications*, vol. 22, Art. no. 100738, 2025, doi: 10.1016/j.mlwa.2025.100738.
[20] N. D. Primadya, A. Nugraha, and S. Y. Fahrezi, "Optimizing imbalanced data classification: Under sampling algorithm strategy with classification combination," *Techné Jurnal Ilmiah Elektroteknika*, pp. 277–288, 2024.
[21] L. Saitta, "Support-vector networks," *Machine Learning*, vol. 297, pp. 273–297, 1995.
[22] N. Galea *et al.*, "Comparative evaluation of Optuna-optimized radial basis function and sigmoid kernels in support vector machine," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 8, no. 3, pp. 677–686, 2025.
[23] F. Genuario *et al.*, "Machine learning-based methodologies for cyber-attacks and network traffic monitoring: A review and insights," *Information*, vol. 15, no. 11, 2024, doi: 10.3390/info15110741.
[24] A. Nanda, H. Wahyu, R. Rahmaddeni, S. Sutisna, and R. Rinaldi, "Perbandingan efektivitas random forest, SVM, dan logistic regression dalam deteksi intrusi jaringan," *JATISI*, vol. 12, no. 2, pp. 129–139, 2025, doi: 10.35957/jatisi.v12i2.10908.
[25] A. Z. K. Matloob, M. I. Kareem, and H. K. Alwan, "Machine learning-based classification models for efficient DDoS detection," *International Journal of Computing and Digital Systems*, vol. 17, no. 1, pp. 1–13, 2025.
[26] M. A. Faizin *et al.*, "Optimizing feature selection method in intrusion detection system using thresholding," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 3, pp. 214–226, 2024, doi: 10.22266/ijies2024.0630.18.
[27] K. Sundaram *et al.*, "A novel hybrid feature selection with cascaded LSTM: Enhancing security in IoT networks," *Wireless Communications and Mobile Computing*, 2024, doi: 10.1155/2024/5522431.
[28] A. Nanda and H. Wahyu Perdana, "Perbandingan efektivitas random forest, SVM, dan logistic regression dalam deteksi intrusi jaringan," vol. 12, no. 2, pp. 129–139, 2025. [Online]. Available: http://jurnal.mdp.ac.id.