IMPROVING INTERNET OF THINGS CYBER ATTACK DETECTION WITH INFORMATION GAIN AND DECISION TREE

PENINGKATAN DETEKSI SERANGAN SIBER PADA INTERNET OF THINGS DENGAN INFORMATION GAIN DAN DECISION TREE

Alvin Mufidha Ahmad[1], Fauzi Adi Rafrastara[2*]
Universitas Dian Nuswantoro, Jl. Nakula I No. 5-11, Semarang
111202013071@mhs.dinus.ac.id[1], fauziadi@dsn.dinus.ac.id[2]*

Abstract *- The rapid proliferation of Internet of Things (IoT) devices within digital ecosystems has enhanced efficiency and availability but has also expanded the attack surface for cyber threats. This study aims to improve intrusion detection accuracy in IoT environments by addressing two key challenges: class imbalance and high feature dimensionality. Random Undersampling (RUS) is employed to mitigate data imbalance in the CIC IoT 2023 dataset, while feature selection is performed using the filter-based Information Gain method. A Decision Tree classifier is implemented and validated using k-fold cross-validation to ensure result reliability. Experimental results demonstrate that the proposed approach achieves an accuracy of 88.7%, outperforming a wrapper-based method, which attained 87.3%. These findings confirm that an appropriately designed filter-based feature selection strategy can effectively enhance the performance of intrusion detection systems for IoT security.*

Keywords *- Machine Learning, Cross validation, Decision tree, Information Gain, Feature Selection.*

Abstrak - Perkembangan pesat perangkat Internet of Things (IoT) dalam ekosistem digital meningkatkan efisiensi dan ketersediaan, tetapi juga memperluas permukaan serangan terhadap ancaman siber. Penelitian ini bertujuan meningkatkan akurasi deteksi intrusi pada perangkat IoT dengan mengatasi dua tantangan utama: ketidakseimbangan kelas dan dimensionalitas tinggi fitur. Pendekatan Random Undersampling (RUS) digunakan untuk menangani ketidakseimbangan data pada dataset CIC IoT 2023, sedangkan seleksi fitur dilakukan menggunakan metode berbasis filter Information Gain. Model klasifikasi dibangun dengan algoritma Decision Tree dan divalidasi menggunakan k-fold cross-validation untuk menjamin keandalan hasil. Eksperimen menunjukkan bahwa pendekatan ini mencapai akurasi sebesar 88,7%, mengungguli metode berbasis wrapper yang hanya mencapai 87,3%. Temuan ini mengonfirmasi bahwa strategi seleksi fitur berbasis filter yang tepat dapat meningkatkan kinerja sistem deteksi intrusi untuk keamanan IoT.

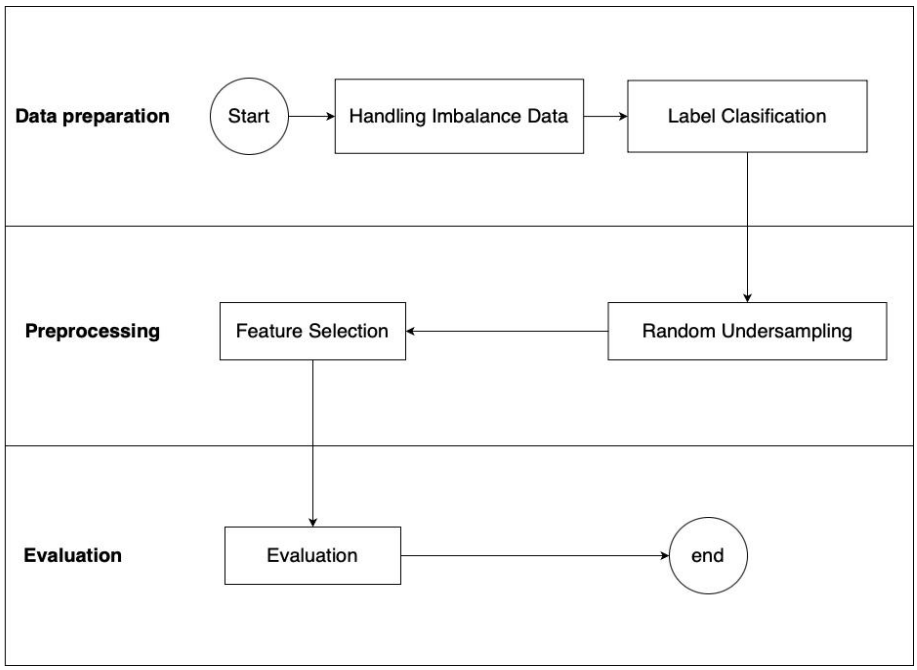
Kata Kunci - Machine Learning, Cross validation, Decision tree, Information Gain, Fitur Seleksi.

I. PENDAHULUAN

Lanskap keamanan siber di tahun 2025 ditandai oleh kemajuan teknologi yang pesat, namun diiringi oleh celah regulasi dan keamanan yang dimanfaatkan penyerang [1]. Sangat penting untuk menyadari bahwa kerentanan keamanan IoT dapat menyebabkan pelanggaran privasi data [2]. Pesatnya pertumbuhan IoT telah menempatkan isu keamanan sebagai perhatian utama [3]. Meningkatnya adopsi perangkat Internet of Things (IoT) telah menghadirkan tantangan baru terkait keamanan dan jaringan. Metode terpusat yang konvensional menghadapi masalah skalabilitas dan kerentanan keamanan seiring pesatnya pertumbuhan perangkat yang saling terhubung [4]. Ancaman terhadap sistem IoT terus berkembang dan beragam, mencakup serangan seperti Denial-of-Service (DoS), Distributed DoS (DDoS), dan reconnaissance. Memahami pola serangan ini dapat didasarkan pada taksonomi serangan siber yang robust, yang berperan crucial dalam mengklasifikasikan dan menganalisis vektor serangan [5]. Dalam situasi digital saat ini, ledakan perangkat Internet of Things (IoT) menghasilkan volume data berdimensi tinggi yang sangat besar, sehingga menimbulkan tantangan signifikan terhadap keamanan data dan privasi [6]. Untuk memitigasi ancaman-ancaman tersebut, pendekatan Machine Learning (ML) banyak diterapkan untuk membangun sistem deteksi intrusi yang cerdas. Di antara berbagai algoritma ML, Decision Tree telah terbukti efektif digunakan untuk tugas deteksi anomali dan intrusi dalam sistem cyber-physical [7]. Berbagai metode seleksi fitur, seperti SHAP, *correlation-based*, dan Information Gain, telah diterapkan dan terbukti secara signifikan meningkatkan kemampuan sistem deteksi intrusi [8]. Model decision tree dengan lebih sedikit variabel dapat berkinerja lebih baik daripada model yang memanfaatkan semua prediktor [9]. Rendahnya akurasi deteksi serangan siber pada sistem IoT pertanian, yang disebabkan oleh ketidakseimbangan kelas dan redundansi fitur pada dataset CIC IoT 2023, menjadi masalah sentral penelitian ini. Kami berhipotesis bahwa penerapan *Random Undersampling* dan seleksi fitur *Information Gain* sebelum *Decision Tree* dapat menghasilkan model yang lebih unggul. Hasil eksperimen tidak hanya mengkonfirmasi hipotesis dengan peningkatan akurasi dan efisiensi, tetapi juga menunjukkan peningkatan ketahanan sistem terhadap ancaman yang relevan seperti DoS, DDoS, dan reconnaissance. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1, yang mencakup tahapan dari akuisisi data, pra-pemrosesan, seleksi fitur, pemodelan, hingga evaluasi.

II. SIGNIFIKANSI STUDI

Penelitian ini mengusulkan pipeline Decision Tree (DT) yang diperkuat Information Gain (IG) dan Random Undersampling (RUS) sebagai baseline yang robust dan dapat diinterpretasi untuk deteksi intrusi IoT. Dalam konteks sistem IoT yang resource-constrained, kompleksitas model yang tinggi tidak selalu terjangkau atau diinginkan [10]. Hal ini memungkinkan ekstraksi fitur-fitur kunci yang terkait dengan risiko dan mendukung pengambilan keputusan yang tepat untuk menangani risiko tersebut [11]. Kombinasi IG dan RUS menciptakan pipeline pra-pemrosesan yang ringan secara komputasi namun efektif, menjawab tantangan data IoT yang tidak seimbang dan berdimensi tinggi. Namun, banyak penelitian sebelumnya sering kali mengabaikan keterbatasan komputasi yang melekat pada perangkat keras IoT, sehingga solusi yang dihasilkan kurang feasible untuk diimplementasikan dalam skenario dunia nyata [12]. Validasi ketat menggunakan 10-fold stratified cross-validation menghasilkan akurasi 88,3% yang stabil. Pendekatan ini diilustrasikan dalam alur kerja penelitian pada Gambar 1. Meskipun tidak melampaui pendekatan ensemble atau deep learning [29]–[32], kontribusi utama terletak pada metodologi yang dapat direplikasi, efisien, dan dapat diinterpretasi, menjembatani kesenjangan antara kompleksitas model dan keterterapan praktis. Berbagai pendekatan pembelajaran mesin telah diterapkan untuk deteksi intrusi IoT dengan hasil yang bervariasi. Tabel 1 merangkum penelitian terkait yang menggunakan berbagai dataset dan algoritma, menunjukkan bahwa pendekatan berbasis Decision Tree pada dataset CIC IoT 2023 masih memiliki ruang untuk peningkatan kinerja.



Gambar 1. Alur Penelitian

A. Sumber Data

Dataset CIC IoT 2023 yang terdiri dari 46.686.579 rekaman dengan 47 fitur awal digunakan dalam penelitian ini. Setelah pembersihan data (missing values = 0; 7.865 duplikat dihapus) dan konversi fitur kategorikal menjadi numerik menggunakan transformasi nominal ke numerik, jumlah fitur meningkat menjadi 61 fitur. Proses ini mengubah variabel kategorikal seperti Protocol_Type menjadi representasi numerik (0 dan 1), yang diperlukan untuk kompatibilitas dengan algoritma Decision Tree. 34 label kelas dikelompokkan menjadi 8 kategori utama (DDoS, DoS, Mirai, Benign, Spoofing, Recon, Web, BruteForce) seperti ditunjukkan pada Tabel 2.

TABEL I
PERBANDINGAN KARYA TERKAIT TENTANG DETEKSI INTRUSI IOT

No	**Research Summary**				
	Penulis	**Dataset**	**Pendekatan Utama**	**Algoritma**	**Kinerja**
1 [29]	Fadil Adji et al. (2023)	CTU-13	Botnet detection berbasis ML	Decision Tree, KNN, Gaussian	Decision Tree: 99.35%, KNN: 99.73%, Gaussian: 88%
2 [30]	Inayah & Ramli (2024)	CIC-ToN-IoT	IDS untuk lingkungan pemerintah	Random Forest	99%,
3 [22]	Setiawan et al. (2024)	CIC loT 2023	Wrapped based feature selection dengan Decision Tree	Decision Tree	87.3 %

4 [31]	Sharma (2024)	CIC IoT 2023	Optimasi CNN dan XGBoost	CNN berbasis fitur statistik	CNN 98.84% dan XGBoost 95.48%
5 [32]	Phan (2024)	CIC IoT 2023	Pendekatan dapat dijelaskan machine learning atau XAI	Random Forest	98.57%

TABEL II
PENGELOMPOKAN 34 LABEL KELAS MENJADI 8 KATEGORI SERANGAN

Kategori	Label Kelas Asli	Jumlah Rekaman
DDOS	DDOS-RSTFINFlood, DDoS-PSHACK_Flood, DDoS-SYN_Flood, DDoS-SlowLoris	33.900.000
DoS	Mirai-greeth_flood, Mirai-greip_flood, Mirai-udpplain	8.100.000
Mirai	BenignTraffic	1.850.000
Benign	DNS_Spoofing, MITM-ArpSpoofing	1.200.000
Spoofing	Recon-PingSweep, Recon-PortScan, Recon-OSScan	850.000
Recon	Sqlinjection, XSS, Uploading_Attack,	450.000
Web	Vulnerability_scanner	290.000
BruteForce	DictionaryBruteForce, BruteForce	13.064

B. Penanganan Ketidakseimbangan Kelas

Mengatasi masalah ketidakseimbangan data merupakan langkah krusial untuk menelaah dampaknya terhadap kinerja IDS dan meningkatkan performa sistem secara keseluruhan [14]. Dalam literatur, teknik oversampling seperti SMOTE beserta varian optimasinya sering menjadi metode utama untuk menghasilkan sampel sintetis guna memperkaya kelas minoritas [13], [15]. Namun, penanganan ketidakseimbangan pada tingkat data (data-level approaches) juga dapat dilakukan melalui undersampling. Berbeda dengan oversampling yang menambah kompleksitas data, metode Random Undersampling (RUS) dipilih dalam penelitian ini karena kontribusinya yang signifikan dalam mengurangi waktu pelatihan dan meningkatkan efisiensi komputasi [16]. Teknik ini menyeimbangkan distribusi data dengan mereduksi sampel kelas mayoritas secara acak agar proporsional. Implementasi RUS dilakukan dengan menyamakan jumlah sampel setiap kelas berdasarkan kelas minoritas (Brute Force, n=13.064), sehingga menghasilkan dataset seimbang dengan total 104.512 rekaman (8 x 13.064).

C. Seleksi Fitur

Untuk memitigasi *noise* dan redundansi pada data berdimensi tinggi, seleksi fitur diterapkan pasca-penyeimbangan data guna mereduksi kompleksitas serta mengoptimalkan akurasi model [17], [18]. Penelitian ini mengadopsi pendekatan *filter-based* karena efisiensi waktu eksekusinya, dengan spesifik menggunakan metode *Information Gain* (IG) yang terbukti efektif dalam mengevaluasi relevansi statistik atribut [19]. Proses reduksi dimensi ini menyaring 46 fitur awal menjadi 10 fitur prioritas dengan skor IG tertinggi untuk pemodelan, yaitu: *flow_duration, Header_Length, Rate, Protocol_Type, ack_flag_number, rst_count, TCP, Min, IAT,* dan *Magnitude.*

D. Pemodelan dengan Decision Tree

Model Decision Tree (CART) dikonstruksi menggunakan kriteria *Gini impurity* dan dievaluasi melalui *Stratified 10-fold Cross-Validation* yang menghasilkan akurasi rata-rata 88,7%. Algoritma ini dipilih karena interpretabilitasnya yang tinggi, efisiensi pada data berskala besar, serta kemampuannya menangani pola non-linear yang terbukti lebih unggul dibandingkan metode seperti Naïve Bayes [20]–[22]. Konfigurasi *hyperparameter* dioptimalkan untuk menyeimbangkan antara kemampuan model menangkap pola detail dan generalisasi:

1. *Unconstrained Depth:* Batasan kedalaman pohon dinonaktifkan untuk memungkinkan model mempelajari struktur hirarkis yang mendalam. Hal ini krusial untuk membedakan kelas serangan dengan kemiripan statistik tinggi, seperti antara *Web Attack* dan *Brute Force*.
2. *Regularization Constraints:* Untuk mencegah *overfitting* akibat kedalaman yang tak terbatas, parameter *minimum instances in leaves* diatur sebesar 20 dan *do not split subsets smaller than* sebesar 100. Batasan ini memastikan bahwa setiap aturan keputusan dan percabangan didukung oleh jumlah sampel yang signifikan secara statistik, sehingga model tidak membentuk pola berdasarkan *noise* atau anomali minor.

E. Evaluasi

Tahap akhir penelitian adalah evaluasi model untuk menilai dan memilih metode klasifikasi terbaik berdasarkan kinerjanya. Kinerja diukur menggunakan confusion matrix, yaitu tabel matriks yang membandingkan hasil prediksi [23]. Metode ini menghasilkan nilai akurasi dalam mengenali objek dari berbagai kelas berbeda. Akurasi digunakan untuk mengevaluasi model klasifikasi dengan menggambarkan proporsi prediksi yang benar dalam sebuah dataset tertentu. Akurasi dihitung menggunakan Persamaan (1) berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} x\ 100\% \quad (1)$$

Dimana:

TP (True Positives): Jumlah prediksi positif yang benar. TN (True Negatives): Jumlah prediksi negatif yang benar. FP (False Positives): Jumlah prediksi positif yang salah. FN (False Negatives): Jumlah data positif yang diprediksi sebagai negatif. Selanjutnya, terdapat Precision, yang mengacu pada tingkat ketepatan antara data yang diminta dengan data yang dikembalikan oleh model [24]. Precision dihitung menggunakan rumus berikut:

$$Precission = \frac{TP}{TP + FP} \quad (2)$$

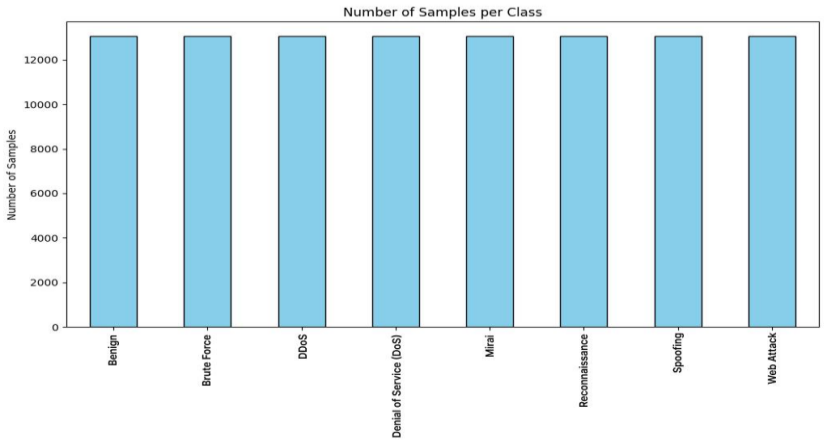
Recall adalah tingkat keberhasilan sebuah sistem dalam mengambil kembali informasi [25]. Recall juga merupakan bagian dari matriks evaluasi dan dihitung berdasarkan Persamaan (3):

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

III. HASIL DAN PEMBAHASAN

A. Preparasi Data dan Penyeimbangan Kelas

Dataset CIC IoT 2023 [26] digunakan, yang mensimulasikan lalu lintas jaringan dari 105 perangkat IoT. Setelah *data cleaning* (Gambar 2), 34 kelas serangan asli dikelompokkan menjadi 8 kategori. Dataset menunjukkan ketidakseimbangan ekstrem, didominasi oleh kelas DDoS (33,9 juta rekaman) dan DoS (8,1 juta), sementara kelas minoritas seperti BruteForce hanya memiliki 13.064 rekaman. Penerapan Random Undersampling (RUS) berhasil menyeimbangkan distribusi, menghasilkan dataset akhir 104.512 rekaman (13.064 per kelas).



Gambar 2. Distribusi Kelas Setelah Pra Pemrosesan

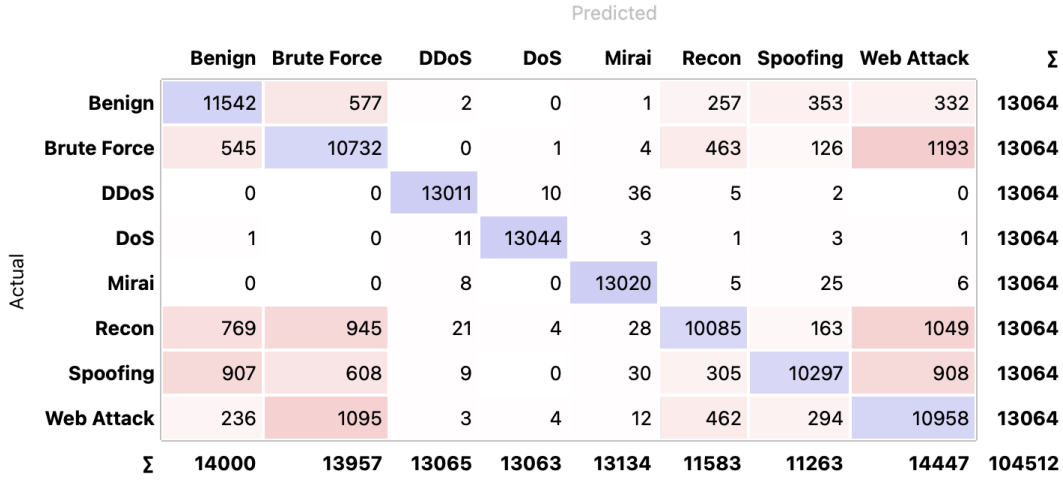
B. Seleksi Fitur dan Pemodelan

Dari 61 fitur setelah preprocessing, Information Gain (IG) dipilih sebagai metode seleksi fitur utama karena efisiensi komputasinya yang tinggi dan kinerja yang stabil dibandingkan metode *wrapper* dan *embedded*. Sepuluh fitur dengan skor IG tertinggi dipilih untuk pemodelan. Evaluasi komparatif lima algoritma Tabel 3 menggunakan *10-fold stratified cross-validation* menunjukan bahwa Decision Tree (DT) dan AdaBoost mencapai kinerja terbaik dengan akurasi 88,5%.

TABEL III
PERBANDINGAN KINERJA LIMA ALGORITMA KLASIFIKASI PADA DATASET CIC IOT 2023

Algoritma	CA	F1	Prec
Decision Tree	0,887	0,887	0,887
AdaBoost	0,88	0,88	0,88
Random Forest	0,87	0,87	0,87
Neural Network	0,75	0,75	0,76
kNN	0,71	0,71	0,72

C. Analisis Kinerja



Gambar 3. Confusion Matrix Model Decision Tree dengan Information Gain pada Dataset CIC IoT 2023

Gambar 3 (Confusion Matrix) menunjukkan detail kinerja klasifikasi model. Model menunjukkan performa hampir sempurna dalam mendeteksi serangan berbasis flooding seperti DoS (99,8%), Mirai (99,6%), dan DDoS (99,6%). Tingginya akurasi pada kelas-kelas ini menunjukkan efektivitas fitur flow_duration dan Rate yang dipilih melalui Information Gain. Namun, tantangan klasifikasi terlihat pada serangan yang memiliki pola trafik serupa. Kesalahan klasifikasi (misclassification) tertinggi terjadi secara silang antara kelas Brute Force dan Web Attack. Sebanyak 9,1% sampel Brute Force salah diprediksi sebagai Web Attack, dan sebaliknya 8,4% Web Attack terdeteksi sebagai Brute Force. Selain itu, kelas Recon juga memiliki tingkat kebingungan yang moderat terhadap Web Attack (8,0%). Hal ini wajar mengingat ketiga jenis serangan tersebut sering kali melibatkan pola permintaan berulang (repetitive requests) yang mirip dalam durasi singkat

D. Perbandingan dengan Studi Terkini dan Signifikansi Kontekstual

Hasil penelitian ini dibandingkan dengan studi Setiawan et al. [22] yang juga menggunakan Decision Tree pada dataset yang sama Tabel 4. Penelitian kami, dengan protokol validasi yang lebih ketat (*10-fold CV*), menghasilkan akurasi yang lebih tinggi dan andal (88,5% vs 87,32%).

TABEL IV
PERBANDINGAN METODOLOGIS: PROTOKOL VALIDASI DAN KINERJA TERHADAP STUDI DASAR [22]

Nomor	Metode	Hasil (dalam persen)	Skema Evaluasi
1	[22] Wrapper dan Decision Tree	87,32	Split 80:20 (tunggal)
2	Penelitian ini: Information Gain dan Decision Tree	88,7	10-fold CV

KESIMPULAN

Penelitian ini mengusulkan model deteksi serangan IoT berbasis Information Gain untuk seleksi fitur dan Decision Tree sebagai klasifikasi, yang dievaluasi menggunakan 10-fold stratified cross-validation pada dataset CIC IoT 2023 dari Canadian Institute for Cybersecurity (CIC), University of New Brunswick. Dataset awal terdiri dari 46.686.579 rekaman dengan 47 fitur kategorikal dan numerik. Setelah pra-pemrosesan yang meliputi pembersihan data, konversi fitur kategorikal menjadi encoding numerik biner (menghasilkan 61 fitur), pengelompokan 34 kelas menjadi 8 kategori, dan penyeimbangan melalui Random Undersampling (menghasilkan 104.512 sampel), dilakukan seleksi 47 fitur terbaik menggunakan Information Gain.

Model Decision Tree yang dibangun mencapai akurasi rata-rata 88,7% (±0,16%) berdasarkan 10-fold stratified cross-validation, dengan precision 89,0%, recall 88,7%, dan F1-score 88,7%. Peningkatan sebesar 1,38 poin persentase dibanding penelitian baseline [22] yang mencapai 87,32% menunjukkan keunggulan metode seleksi fitur berbasis filter yang diusulkan. Hasil ini menjadi lebih signifikan mengingat dicapai melalui skema validasi yang lebih ketat (10-fold cross-validation) dibandingkan baseline (single train-test split), yang mengonfirmasi bahwa model tidak hanya lebih akurat tetapi juga lebih andal (robust) dalam menangani variasi data.

Kontribusi utama penelitian ini adalah:

1. Demonstrasi bahwa kombinasi Information Gain dan Random Undersampling dapat menghasilkan model deteksi intrusi IoT yang akurat dengan kompleksitas rendah
2. Validasi ketat menggunakan 10-fold stratified CV yang memastikan generalisasi model
3. Identifikasi 10 fitur optimal yang dapat diimplementasikan pada perangkat IoT dengan resource terbatas

Keterbatasan dan penelitian lanjutan: Meskipun model menunjukkan kinerja baik pada kelas mayoritas (DDoS, DoS), masih terdapat ruang perbaikan pada deteksi kelas Mirai dan Spoofing yang memiliki confusion rate lebih tinggi. Penelitian selanjutnya dapat mengeksplorasi:

- Teknik feature engineering khusus untuk membedakan Mirai dari DDoS
- Ensemble methods untuk meningkatkan deteksi pada kelas yang overlapping
- Implementasi real-time pada perangkat IoT embedded untuk validasi praktik.

REFERENSI

[1] A. El-Deeb, "Are we safe online? Cybersecurity as the golden word in 2025," ACM SIGSOFT Softw. Eng. Notes, vol. 50, no. 2, pp. 10–10, 2025.

[2] N. Sharma and P. Dhiman, "A survey on IoT security: Challenges and their solutions using machine learning and blockchain technology," Cluster Comput., vol. 28, no. 5, p. 313, 2025.

[3] N. K. Upadhyay, S. Periyasamy, and V. Kumar, "Lightweight cryptographic protocols for enhanced security in the IoT: A survey," in Proc. 3rd Int. Conf. Artif. Intell. Mach. Learn. Appl., 2025, pp. 1–6.

[4] V. Maurya et al., "Blockchain-driven security for IoT networks: State-of-the-art, challenges and future directions," Peer-to-Peer Netw. Appl., vol. 18, no. 1, p. 53, 2025.

[5] P. R. da P. F. Santos et al., "Towards robust cyber attack taxonomies: A survey with requirements, structures, and assessment," ACM Comput. Surv., vol. 57, no. 8, pp. 1–36, 2025.

[6] F. S. Alrayes et al., "Privacy-preserving approach for IoT networks using statistical learning with optimization algorithm on high-dimensional big data environment," Sci. Rep., vol. 15, no. 1, p. 3338, 2025.

[7] A. Samiah, M. A. Umer, and S. Siddiqui, "Decision tree based invariants for intrusion detection in industrial control system," Comput. Secur., p. 104511, 2025.
[8] A. S. Anagha, C. Thomas, and N. Balakrishnan, "Optimized intrusion predictions through feature selection methods," *Comput. Secur.*, p. 104541, 2025.
[9] H. Al-Najjar *et al.*, "Decision tree-based IoT botnet attack detection," in *Proc. 2nd Int. Conf. Cyber Resilience (ICCR)*, 2024, pp. 1–5.
[10] A. Wakili and S. Bakkali, "Privacy-preserving security of IoT networks: A comparative analysis of methods and applications," Cyber Secur. Appl., vol. 3, p. 100084, 2025.
[11] S. Islam et al., "Intelligent dynamic cybersecurity risk management framework with explainability and interpretability of AI models," J. Reliable Intell. Environ., vol. 11, no. 3, p. 12, 2025.
[12] N. Albanbay et al., "Federated learning-based intrusion detection in IoT networks: Performance evaluation and data scaling study," J. Sens. Actuator Netw., vol. 14, no. 4, p. 78, 2025.
[13] B. Gomathy et al., "Optimized SMOTE for imbalanced data handling in machine learning," in Proc. 3rd Int. Conf. Adv. Comput. Comput. Technol., 2025, pp. 1–5.
[14] Y. Zhang, R. C. Muniyandi, and F. Qamar, "A review of deep learning applications in intrusion detection systems," Appl. Sci., vol. 15, no. 3, p. 1552, 2025.
[15] G. Zhao et al., "LGSMOTE-IDS: Line graph based weighted-distance SMOTE for imbalanced network traffic detection," Expert Syst. Appl., p. 127645, 2025.
[16] M. D. Firmansyah, I. Rizqa, and F. A. Rafrastara, "Balancing CICIoV2024 dataset with RUS for improved IoV attack detection," J. Appl. Inform. Comput., vol. 9, no. 2, pp. 250–257, 2025.
[17] E. N. R. Khakim, A. Hermawan, and D. Avianto, "Implementasi correlation matrix pada klasifikasi dataset wine," JIKO, vol. 7, no. 1, pp. 158–166, 2023.
[18] N. T. Romadloni and N. D. Septiyanti, "Optimasi feature selection pada komentar media sosial terhadap peralihan TV digital," Decode, vol. 3, no. 2, pp. 151–160, 2023.
[19] A. Devia and B. Soewito, "Analisis perbandingan metode seleksi fitur untuk mendeteksi anomali pada dataset CIC-IDS-2018," J. Teknol. Sist. Inform. Bisnis, vol. 5, no. 4, pp. 572–578, 2023.
[20] T. Becker et al., "Decision trees and random forests," Am. J. Orthod. Dentofacial Orthop., vol. 164, no. 6, pp. 894–897, 2023.
[21] Sulistyaningrum et al., "Penanganan missing value dan perbandingan performa naïve bayes dan decision tree," ALMUISY, vol. 4, no. 1, pp. 21–26, 2025.
[22] D. Setiawan, A. Nugraha, and A. Luthfiarta, "Komparasi teknik feature selection dalam klasifikasi serangan IoT menggunakan decision tree," *J. Media Inform. Budidarma*, vol. 8, no. 1, pp. 83–93, 2024.
[23] F. R. Valerian, M. Syarief, and D. A. Fatah, "Klasifikasi tingkat obesitas menggunakan metode GBM dan confusion matrix," JATI, vol. 9, no. 2, pp. 2242–2249, 2025
[24] N. J. Hayati, D. Singasatia, and M. R. Muttaqin, "Object tracking menggunakan YOLOv8 untuk menghitung kendaraan," Komputa, vol. 12, no. 2, pp. 91–99, 2023.
[25] E. Utami, "Comparison of naïve Bayes, KNN, and SVM in Twitter user classification," JUTIF, vol. 3, no. 3, pp. 673–680, 2022.
[26] E. C. P. Neto et al., "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," Sensors, vol. 23, no. 13, p. 5941, 2023.
[27] P. Modi, "Towards efficient machine learning method for IoT DDoS attack detection," arXiv preprint arXiv:2408.10267, 2024.
[28] A. I. Jony and A. K. B. Arnob, "A long short-term memory based approach for detecting cyber attacks in IoT using CIC IoT2023 dataset," J. Edge Comput., vol. 3, no. 1, pp. 28–42, 2024.
[29] R. F. A. Bintoro, P. H. Trisnawan, and M. Data, "Deteksi botnet menggunakan metode decision tree pada dataset CTU," JPTIIK, vol. 7, no. 6, pp. 2921–2930, 2023.
[30] K. Inayah and K. Ramli, "Analisis kinerja IDS berbasis random forest menggunakan dataset unbalanced honeynet BSSN," JPTIIK, vol. 11, no. 4, pp. 867–876, 2024.
[31] A. Sharma, H. Babbar, and A. K. Vats, "Detecting reconnaissance activities in IoT networks using UNB CIC IoT 2023 dataset," in Proc. 4th ASIANCON, 2024, pp. 1–6.
[32] V. A. Phan, J. Jerabek, and L. Malina, "Comparison of multiple feature selection techniques for ML-based detection of IoT attacks," in *Proc. 19th Int. Conf. ARES*, 2024, pp. 1–8.