# Deteksi Serangan Brute Force pada Jaringan IoT Menggunakan Algoritma Machine Learning Berbasis Dataset CICIoT2023

Rizqi Bhamaryuda
Fakultas Informatika
Universitas Telkom
Bandung
rizqibryda@student.telkomuniversity.ac.id

Aji Gautama Putrada
Fakultas Informatika
Universitas Telkom
Bandung
ajigps@telkomuniversity.ac.id

Febri Dawani
Fakultas Informatika
Universitas Telkom
Bandung
febridawani@telkomuniversity.ac.id

**Abstrak**

**Perkembangan Internet of Things (IoT) yang dipakai banyak orang terus digunakan dengan berbagai perkembangan dan penemuan baru. Dalam proses perkembangannya dan waktu, seiringnya juga dengan munculnya tantangan yang dapat terjadi perangkat IoT dengan berbagai macam permasalahan. Salah satu kejadian dalam permasalahan IoT adalah serangan *brute force* yang merupakan upaya berbagai gabungan kredensial seperti kata sandi yang dilakukan untuk memaksa masuk ke akses perangkatnya hingga menemukan akses yang sah secara brutal atau banyak sekali percobaannya. Dari aktivitas tersebut dikumpulkan menjadi dataset CICIoT2023 yang dilakukan untuk mengenali aktivitas tersebut sebagai sistem deteksi intrusi (IDS) sehingga harapan kedepannya dapat memantau dan mencegah kerusakan yang tidak diinginkan dengan melakukan pembersihan data dan penerapan *machine learning*. Hasil perbandingan antara *Random Forest* dan *Decision Tree* menunjukkan bahwa akurasi mencapai 0.9466 dengan algoritma model Random Forest dengan teknik *Random Undersampling* (RUS) sehingga ini menunjukkan bahwa keefektivitas dari metode ensemble yang ada dalam meningkatkan stabilitas deteksi anomali serangan brute force pada lingkungan IoT.**

**Kata kunci: Internet of Things, Machine Learning, Cyber Security, CICIoT2023, Intrusion Detection System**

**Abstract**

**The development of IoT, which is widely used by many people, continues to be used with various developments and new discoveries. As the process of development and time goes by, challenges also arise that can occur with various problems in IoT devices. One incident in the IoT problem is a brute force attack, which is an attempt to combine various credentials such as passwords that are carried out to force entry into the device until finding legitimate access by brute force or numerous attempts. From these activities collected into the CICIoT2023 dataset which is carried out to recognize these activities as an intrusion detection system (IDS) so that in the future it is hoped that it can monitor and prevent unwanted damage by performing data cleaning and applying machine learning. The results show that the accuracy reaches 0.9466 with the Random Forest model algorithm with the Random Undersampling (RUS) technique, thus demonstrating the effectiveness of the existing ensemble method in improving the stability of anomaly detection of brute force attacks in the IoT environment.**

**Keywords: Internet of Things, Machine Learning, Cyber Security, CICIoT2023, Intrusion Detection System**

## 1. Pendahuluan

### Latar Belakang

Internet of Things (IoT) dapat digunakan berbagai tempat, seperti di rumah, tempat umum, parkiran, dan sebagainya yang digunakan untuk memantau mengendalikan, atau hal lainnya yang bisa dilakukan dengannya sehingga IoT telah menarik banyak perhatian sehingga banyak dikembangkan dan dipergunakan seperti kepada industri, militer, logistik, transportasi, kesehatan, tenaga listrik jaringan, pertanian, energi konservasi, perlindungan lingkungan, perawatan kesehatan, *smart homes*, dan sebagainya [1], [2].

Dalam beberapa tahun terakhir ini, beberapa peneliti melakukan penelitian privasi, perlindungan perangkat dan berdiskusi mengenai hukumnya kerangka kerja perlindungan privasi mengenai IoT [1]. Hal ini diperkuat melalui dalam paper [3] bahwa data *crowdsourcing* IoT menyebabkan ancaman privasi dan keamanan yang tidak terprediksi sehingga perlu dilindungi untuk beroperasi di lingkungan dunia nyata. IoT dapat didefinisikan sebagai jaringan yang berinteraksi dan terhubung secara penuh antar sosial dan benda ke benda yang melalui peralatan penginderaan informasi sesuai dengan protokol yang disepakati [4], [5].

Setiap inovasi teknologi yang dapat mengintegrasikan sesuatu ini pasti menimbulkan tantangannya sendiri, seperti kebocoran data, mengambil kendali tanpa izin atau dilakukan oleh pihak yang tidak bertanggung jawab, dan sebagainya. Ada berbagai cara untuk bisa menyalahgunakan kerentanan pada perangkat IoT, salah satunya adalah brute force, yaitu teknik penyerangan yang mencoba berbagai kemungkinan kombinasi kredensial secara sistematis [5], [6].

Mengingat skala serangan otomatis dengan berbagai macam respons ini dibutuhkan efektivitas dengan cara melakukan pra-pemrosesan data berbagai sampling dengan undersampling dan oversampling [6], [7]. Kemudian, penerapan machine learning muncul sebagai solusi yang mampu beradaptasi dan merespons ancaman yang kompleks dan terus berkembang di lingkungan IoT dengan menggunakan Decision Tree dipilih sebagai model klasifikasi yang digunakan atas efisiensinya dalam memetakan pola lalu lintas jaringan ke dalam struktur hirarki yang mudah diimplementasikan untuk memahami logika deteksi serangannya. Random Forest dipilih sebagai metode *ensemble* untuk meningkatkan stabilitas deteksi dan mengatasi kelemahannya, seperti risiko *overfitting* sehingga dapat memberikan performa yang lebih konsisten. IDS berbasis machine learning dapat belajar dari data historis untuk mendeteksi perilaku abnormal sehingga menawarkan kemampuan yang lebih baik untuk mengidentifikasi serangan baru yang belum pernah dikenali sebelumnya.

Dengan menganalisis berbagai fitur dari lalu lintas jaringan, model machine learning dapat dilatih untuk mempelajari dan mengenali pola-pola perilaku yang mencurigakan yang menjadi ciri khas dari serangan brute force [6], [8], [9].

**Topik dan Batasannya**

Untuk memfokuskan penelitian pada tujuan yang sudah ditentukan, batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset tersebut hanya digunakan hanya untuk mendeteksi brute force dan aktivitas normal.
2. Penerapan teknis data preprocessing yang digunakan di sini hanya Tomek Links, Near Miss, SMOTE, ADASYN, ENN, dan Random Undersampling.
3. Penerapan machine learning yang digunakan, yaitu Decision Tree dan Random Forest.

**Tujuan**

Berdasarkan perumusan masalah yang telah dirumuskan, maka tujuan yang ingin dicapai pada penelitian ini adalah:

1. Menerapkan teknik sampling Random Undersampling, Near Miss, ENN, dan Tomek Links sebagai kategori undersampling dengan cara mengurangi jumlah sampel pada kelas mayoritasnya. SMOTE dan ADASYN sebagai kategori oversampling dengan cara mengurangi jumlah sampel pada kelas mayoritasnya. Secara umum, berfungsi untuk mengatasi ketidakseimbangan kelas dan menghasilkan dataset latih yang seimbang.
2. Merancang machine learning decision tree dan random forest untuk membedakan antar lalu lintas BenignTraffic sebagai aktivitas normal dan menggunakan DictionaryBruteForce sebagai serangan.
3. Menganalisis dan membandingkan antara dua machine learning dari segi recall.

## 2. Studi Terkait

Pada penelitian lain yang dilakukan oleh El Hariri et al. pada paper "Improving Deep Learning Performance Using Sampling Techniques for IoT Imbalanced Data" [7] menjelaskan bahwa data IoT bisa terdeteksi instruksi dengan beberapa teknik penyeimbangan data, yaitu menggunakan SMOTE (Synthetic Minority Oversampling Technique), ADASYN (Adaptive Synthetic), dan RUS (Random Under Sampling). Kemudian, melakukan penerapan Deep Learning untuk mendeteksi instrusi setelah penyeimbangan data dengan LSTM (Long Short-Term Memory), CNN (Convolutional Neural Networks), Hybrid CNN-LSTM. Kinerja yang terbagus dan luar biasa dengan mencapai hasilnya dari akurasi dan presisi 99.49%, dan dari AUC mencapai 99.93%.

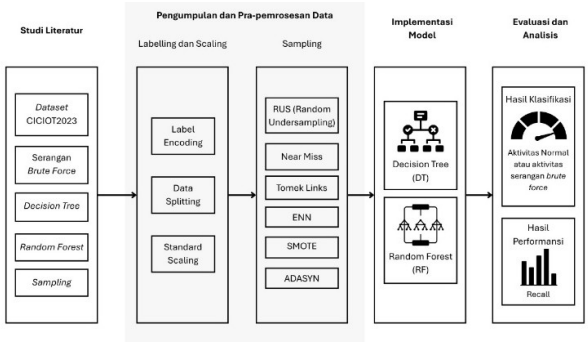
Penelitian lainnya dilakukan oleh Altulaihan et al. pada paper "Anomaly Detection IDS for Detecting DoS Attacks in IoT Networks Based on Machine Learning Algorithms" [9] bahwa pada peningkatnya serangan Denial of Service (DoS) maka upaya yang lakukan adalah dengan melakukan anomaly detection dan algoritma klasifikasi supervised dengan empat jenis, yaitu Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor, dan Support Vector Machine (SVM) pada dataset IoTID20. Kinerja terbaik dalam mendeteksi DoS adalah model DT dan RF yang dilatih dengan fitur-fitur yang telah dipilih menggunakan GA (Genetic Algorithm).

Penelitian lainnya yang telah dilakukan oleh "Random Kernel Forests" [10] perbandingan dari algoritma Random Kernel Forests dengan Random Forest, CO2 Forest, OC1 (Oblique Decision Trees), dan WODT (Weighted Oblique Decision Tree) dengan beberapa dataset nonIoT yang dilakukan untuk mengukur performa akurasi, presisi, dan recall.

## 3. Sistem yang Dibangun

### 3.1 Studi Literatur

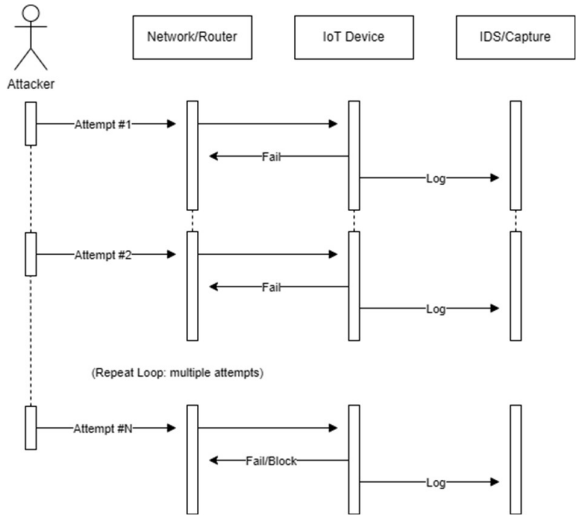
Alur yang disusun untuk sebagai penggambaran prosesnya yang dilakukan untuk mencapai tujuan penelitian secara terstruktur dan sistematis. Alur penelitian dapat dilihat pada diagram berikut:




Gambar 1 Perancangan Sistem

#### 3.1.1 Identifikasi Masalah

Pada tahap awal ini, dilakukan identifikasi terhadap masalah yang melatarbelakangi penelitian. Masalah yang diangkat adalah tingginya kerentanan perangkat IoT terhadap serangan brute force yang memanfaatkan lemahnya sistem otentikasi.




Gambar 2 Proses Brute Force

Pada Gambar 2, terdapat diagram yang menjelaskan alur kerja penyerang melakukan brute force yang mengawali serangan dengan kombinasi kredensial pertama untuk mendapatkan akses perangkat IoT. Input tersebut akan dikirimkan melalui jaringan dan perangkat IoT untuk diverifikasi, upaya otentikasi tersebut gagal dan tercatat oleh sistem. Selanjutnya, penyerang mengulangi dengan kombinasi kredensial yang berbeda dengan hasil gagal dan aktivitasnya direkam. Untuk seterusnya, penyerang melakukan input dengan berbagai kombinasi kredensial

yang berbeda sehingga mengakibatkan kegagalan dan hingga dapat diblokir oleh sistem. Dari aktivitas yang telah terekam ini, terbentuklah menjadi dataset.

3.1.2Studi Literatur

Pada bagian ini, akan ditelusuri lebih lanjut mengenai dataset CICIoT2023 dan penelitian yang telah dilakukan sebelumnya dengan dataset yang sama. Hal ini bertujuan untuk membandingkan hasil dan mengidentifikasi permasalahan yang relevan dengan topik penelitian ini.

3.1.3Pengumpulan Data

Dataset CICIoT2023 yang digunakan untuk diteliti yang kemudian akan dilatih dengan diawali dengan melakukan pra-pemrosesan data, mengimplementasikan data, dan dianalisis.

3.1.4Pra-pemrosesan Data

Tahap ini akan dilakukannya pra-pemrosesan dengan melakukan label encoding, standard scalling, lalu dilanjutkan penerapan sampling Tomek Links, Near Miss, SMOTE, ADASYN, ENN, dan Random Undersampling.

3.1.5Implementasi Model

Pengimplementasian Model akan dilakukannya dengan decision tree dan random forest.

3.1.6Evaluasi dan Analisis

Tahap terakhir mengukur performa akurasi, presisi, dan recall sehingga dapat dibandingkan dari model machine learning yang digunakan performa yang bagus.

3.2 Pra-pemrosesan Data

Tahap ini bertujuan menghasilkan dataset yang siap digunakan untuk dimplementasikan model machine learning.

3.2.1 Random Undersampling

Pra-pemrosesan ini melakukan secara acak memilih sampel dari kelas mayoritas dan menghapusnya dari dataset training [11].

3.2.2 Tomek Links

Teknik yang berfungsi untuk mengatasi ketidakseimbangan kelas (class imbalance) sehingga menghasilkan distribusi data yang lebih seimbang [12].

3.2.3 Near Miss

Teknik yang digunakan untuk menstabilkan distribusi kelas dengan menghapus secara acak contoh-contoh kelas mayoritas sehingga menghasilkan jumlah catatan yang sama untuk kedua kelas [13].

3.2.4 Synthetic Minority Over-sampling Technique (SMOTE)

Teknik yang merancang mengatasi masalah ketidakseimbangan dataset dengan menciptakan sampel sintetis dari kelas minoritas [7].

$$\boldsymbol{x}_{\{new\}} = \boldsymbol{x}_i + \boldsymbol{\lambda}\left(\boldsymbol{x}_{\{neighbor\}} - \boldsymbol{x}_i\right) \quad (1)$$

3.2.5 Adaptive Synthetic Sampling Approach (ADASYN)

Teknik oversampling yang dirancang untuk menyeimbangkan dataset dengan menciptakan dan menambahkan data point sintetis ke kelas minoritas [7].

3.2.6 Edited Nearest Neighbors (ENN)

Metode sebagai pembanding dari Shuffle Split-Edited Nearest Neighbor (SSENN) dengan cara kerjanya adalah mengurangi data latih (training data) dengan menghilangkan data yang salah diklasifikasikan menggunakan metode KNN [14].

3.3 Perancangan Model Klasifikasi

3.3.1Decision Tree (DT)

Algoritma machine learning terawasi berguna untuk masalah klasifikasi dengan cara memecah data yang lebih besar menjadi bagian-bagian yang lebih kecil untuk semua datanya [6]. Proses DT ini akan dimulai dari akar pohon yang akan bercabang menciptakan struktur seperti pohon. Dalam artian, data akan terbagi menjadi dua atau lebih untuk membagi data pelatihan menjadi subset-subset berdasarkan nilai fitur dan tujuannya adalah untuk meminimalkan kriteria ketidakmurnian (*impurity*) [9], [10].

3.3.2Random Forest (RF)

Random Forest merupakan algoritma machine learning terawasi yang digunakan untuk memecahkan masalah klasifikasi dengan kemampuannya yang dapat memberikan prediksi akhir yang lebih akurat dan stabil dibandingkan dengan menggunakan satu model tunggal [9], [15].

3.3.3Skenario Pengujian dan Evaluasi

Di antara setiap tahapan pra-pemrosesan data dapat dilakukan dengan cara memperlakukannya untuk dibandingkan dari setiap proses metode penyimbangan dataset yang paling unggul. Hal ini dilakukan dari dataset yang tidak seimbang sehingga dilakukan pra-pemrosesannya agar model yang akan dipakai menjadi efisiensi, objektif, dan efektif untuk meningkatkan kualitas data. Model machine learning dapat digunakan setelah melakukan pembersihan dan siap dipakai untuk mendeteksi aktivitas anomali, terkhususnya mencari brute force sehingga dapat dikenali aktivitasnya agar kejadian yang sama kembali terjadi yang akan datang, ia dapat terlatihkan untuk melakukan deteksi dan menerapkan solusi atas masalah tersebut [9].

## 4. Evaluasi

Bagian ini berisi dua sub-bagian, yaitu Hasil Pengujian dan Analisis Hasil Pengujian. Pengujian dan analisis yang dilakukan selaras dengan tujuan TA sebagaimana dinyatakan dalam Pendahuluan.

4.1 Skenario Percobaan

Pengujian dilakukan dengan menampilkan hasil utama terlebih dahulu yang kemudian diikuti oleh rincian hasil yang lebih mendalam untuk memberikan gambaran performa secara komprehensif. Mengingat kualitas hasil pengujian bersifat relatif dan menyertakan algoritma pembanding yang kemudian dijalankan pada lingkungan dan dataset yang sama agar memastikan evaluasi yang objektif dan seluruh datanya disajikan melalui tabel dan

diagram yang relevan agar mempermudah perbandingan antar pendekatan yang diuji.

### 4.2 Analisis Hasil Pengujian

Berdasarkan hasil analisis terhadap eksperimen deteksi serangan brute force pada jaringan IoT, diketahui bahwa penerapan teknik sampling telah berhasil menyeimbangkan distribusi kelas pada dataset CICIoT2023 dengan variasi jumlah sampel antara 2.466 hingga 207.260. Keberhasilan penyeimbangan data tersebut didukung oleh rancangan model machine learning menggunakan Decision Tree dan Random Forest yang mampu mengklasifikasikan lalu lintas normal serta serangan brute force secara akurat. Secara spesifik, perbandingan performa menunjukkan keunggulan signifikan Random Forest dibandingkan Decision Tree, terutama melalui penggunaan teknik Random Undersampling yang menghasilkan nilai performa sebesar 0,9466, sekaligus membuktikan efektivitas metode ensemble dalam meningkatkan stabilitas deteksi anomali pada lingkungan IoT.

## 5. Kesimpulan

Berdasarkan hasil analisis terhadap eksperimen deteksi serangan brute force pada jaringan IoT, dapat disimpulkan bahwa penerapan teknik sampling telah berhasil menyeimbangkan distribusi kelas pada dataset CICIoT2023 dengan variasi jumlah sampel antara 2.466 hingga 207.260. Keberhasilan penyeimbangan data tersebut dioptimalkan melalui rancangan model machine learning menggunakan Decision Tree dan Random Forest yang mampu mengklasifikasikan lalu lintas normal serta serangan brute force secara akurat. Secara spesifik, hasil perbandingan menunjukkan keunggulan Random Forest dibandingkan Decision Tree, terutama ketika dikombinasikan dengan teknik Random Undersampling yang menghasilkan performa sebesar 0,9466, sekaligus membuktikan bahwa metode ensemble sangat efektif dalam meningkatkan stabilitas deteksi anomali pada lingkungan IoT.

## REFERENSI


[1] P. Sun, S. Shen, Y. Wan, Z. Wu, Z. Fang, and X. Z. Gao, “A Survey of IoT Privacy Security: Architecture, Technology, Challenges, and Trends,” *IEEE Internet Things J.*, vol. 11, no. 21, pp. 34567–34591, 2024, doi: 10.1109/JIOT.2024.3372518.

[2] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, “CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment,” *Sensors*, vol. 23, no. 13, Jul. 2023, doi: 10.3390/s23135941.

[3] S. Sodagari, “Trends for Mobile IoT Crowdsourcing Privacy and Security in the Big Data Era,” *IEEE Transactions on Technology and Society*, vol. 3, no. 3, pp. 199–225, Jul. 2022, doi: 10.1109/TTS.2022.3191515.

[4] M. Bouzidi, N. Gupta, F. A. Cheikh, A. Shalaginov, and M. Derawi, “A Novel Architectural Framework on IoT Ecosystem, Security Aspects and Mechanisms: A Comprehensive Survey,” *IEEE Access*, vol. 10, pp. 101362–101384, 2022, doi: 10.1109/ACCESS.2022.3207472.

[5] M. Aziz Al Kabir, W. Elmedany, and M. S. Sharif, “Securing IoT Devices Against Emerging Security Threats: Challenges and Mitigation Techniques,” 2023, *Taylor and Francis Ltd.* doi: 10.1080/23742917.2023.2228053.

[6] W. A. H. Salman and C. H. Yong, “Overview of the CICIoT2023 Dataset for Internet of Things Intrusion Detection Systems,” *Mesopotamian Journal of Big Data*, vol. 2025, pp. 50–60, Jun. 2025, doi: 10.58496/MJBD/2025/004.

[7] A. El Hariri, M. Mouiti, O. Habibi, and M. Lazaar, “Improving Deep Learning Performance Using Sampling Techniques for IoT Imbalanced Data,” in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 180–187. doi: 10.1016/j.procs.2023.09.026.

[8] A. Wakili and S. Bakkali, “A resilient IoT intrusion detection system using hybrid feature selection and explainable ensemble learning,” *Results in Engineering*, vol. 28, Dec. 2025, doi: 10.1016/j.rineng.2025.107392.

[9] E. Altulaihan, M. A. Almaiah, and A. Aljughaiman, “Anomaly Detection IDS for Detecting DoS Attacks in IoT Networks Based on Machine Learning Algorithms,” *Sensors*, vol. 24, no. 2, p. 713, Jan. 2024, doi: 10.3390/s24020713.

[10] D. A. Devyatkin and O. G. Grigoriev, “Random Kernel Forests,” *IEEE Access*, vol. 10, pp. 77962–77979, Nov. 2022, doi: 10.1109/ACCESS.2022.3193385.

[11] Z. Fan, S. Sohail, F. Sabrina, and X. Gu, “Sampling-Based Machine Learning Models for Intrusion Detection in Imbalanced Dataset,” *Electronics 2024, Vol. 13, Page 1878*, vol. 13, no. 10, p. 1878, May 2024, doi: 10.3390/ELECTRONICS13101878.

[12] I. S. Thaseen *et al.*, “A Hadoop Based Framework Integrating Machine Learning Classifiers for Anomaly Detection in the Internet of Things,” *Electronics 2021, Vol. 10,* vol. 10, no. 16, Aug. 2021, doi: 10.3390/ELECTRONICS10161955.

[13] N. M. Mqadi, N. Naicker, and T. Adeliyi, “Solving Misclassification of the Credit Card Imbalance Problem Using Near Miss,” *Math. Probl. Eng.*, vol. 2021, no. 1, p. 7194728, Jan. 2021, doi: 10.1155/2021/7194728.

[14] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, “Shuffle Split-Edited Nearest Neighbor: A Novel Intelligent Control Model Compression for Smart Lighting in Edge Computing Environment,” *Smart Innovation, Systems and Technologies*, vol. 324, pp. 219–227, 2023, doi: 10.1007/978-981-19-7447-2_20.

[15] D.-H. Jeong *et al.*, “A Comparative Study on the Influence of Undersampling and Oversampling Techniques for the Classification of Physical Activities Using an Imbalanced Accelerometer Dataset,” *Healthcare 2022, Vol. 10,* vol. 10, no. 7, Jul. 2022, doi: 10.3390/HEALTHCARE10071255.