

## DETEksi ANOMALI PADA JARINGAN INTERNET OF THINGS (IoT) MENGGUNAKAN ALGORITMA RANDOM FOREST DENGAN DATASET RT-IOT2022

Aisyah

Department of Information System  
STMIK Tazkia  
Bogor, Indonesia  
[Aisyahnasution3300@gmail.com](mailto:Aisyahnasution3300@gmail.com)

Prameswari Kirana Jingga

Department of Information System  
STMIK Tazkia  
Bogor, Indonesia  
[Kiranajingga631@gmail.com](mailto:Kiranajingga631@gmail.com)

### Abstrak

Penelitian ini membahas penerapan algoritma machine learning Random Forest dalam proses deteksi anomali pada jaringan Internet of Things (IoT) menggunakan dataset RT-IoT2022. Peningkatan jumlah perangkat IoT menyebabkan tingginya tingkat risiko serangan siber seperti Distributed Denial of Service (DDoS), brute force, port scanning, dan penyalahgunaan protokol. Oleh karena itu, diperlukan metode yang mampu mengidentifikasi aktivitas jaringan yang tidak normal secara akurat. Penelitian dilakukan melalui beberapa tahapan, mulai dari preprocessing data, encoding, scaling, pembagian data, pelatihan model, hingga evaluasi performa. Dataset RT-IoT2022 yang digunakan terdiri dari 123.117 baris dan 85 fitur, kemudian dibersihkan sehingga tersisa 23.839 baris data. Hasil penelitian menunjukkan bahwa Random Forest mampu mendeteksi anomali dengan akurasi sebesar 98,74%, presisi 98,73%, recall 98,74%, dan F1-score 98,73%. Selain itu, fitur seperti flow\_iat.min, fwd\_PSH\_flag\_count, dan fwd\_pkts\_payload.avg menjadi fitur yang paling berpengaruh dalam menentukan anomali. Hasil ini membuktikan bahwa Random Forest efektif digunakan sebagai model deteksi intrusi pada lingkungan IoT.

**Kata kunci :** IoT, Deteksi Anomali, Random Forest, Keamanan Jaringan, Machine Learning

## ANOMALY DETECTION IN INTERNET OF THINGS (IOT) NETWORK WITH RT-IOT2022 DATASET USING RANDOM FOREST

### Abstract

This research discusses the application of the Random Forest machine learning algorithm for anomaly detection in Internet of Things (IoT) networks using the RT-IoT2022 dataset. The increasing number of IoT devices has led to a higher risk of cyberattacks such as Distributed Denial of Service (DDoS), brute force, port scanning, and protocol misuse. Therefore, a method capable of accurately identifying abnormal network activity is required. The research was conducted through several stages, including data preprocessing, encoding, scaling, data splitting, model training, and performance evaluation. The RT-IoT2022 dataset used consists of 123,117 rows and 85 features, and after cleaning, 23,839 rows remained. The results show that Random Forest successfully detects anomalies with an accuracy of 98.74%, precision of 98.73%, recall of 98.74%, and an F1-score of 98.73%. Moreover, features such as flow\_iat.min,

fwd\_PSH\_flag\_count, and fwd\_pkts\_payload.avg became the most influential attributes in determining anomalies. These findings indicate that Random Forest is effective as an intrusion detection model in IoT environments.

**Keywords :** *IoT, anomaly detection, Random Forest, cybersecurity, machine learning*

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Internet of Things (IoT) merupakan teknologi yang memungkinkan perangkat saling terhubung dan bertukar data secara otomatis. Pemanfaatan IoT yang semakin luas menimbulkan tantangan keamanan yang signifikan, terutama karena perangkat IoT memiliki kemampuan komputasi yang terbatas dan seringkali tidak dilengkapi sistem keamanan memadai [1]. Serangan yang umum terjadi pada jaringan IoT meliputi Distributed Denial of Service (DDoS), password attack, port scanning, serta penyalahgunaan protokol seperti MQTT [2][3]. Deteksi anomali digunakan untuk mengidentifikasi aktivitas jaringan yang tidak normal berdasarkan pola data. Pendekatan berbasis machine learning, khususnya Random Forest, banyak digunakan karena mampu menghasilkan akurasi tinggi dan memberikan analisis fitur penting [4]. Dataset RT-IoT2022 digunakan dalam penelitian ini karena menyediakan trafik IoT yang mencakup aktivitas normal dan serangan.

### 1.2 Tinjauan Literatur

Beberapa penelitian sebelumnya telah mengeksplorasi deteksi anomali di IoT menggunakan machine learning. Misalnya, Ferrag et al. [1] membahas deep learning untuk deteksi intrusi, sementara Vaidya et al. [2] melakukan survei teknik anomali di IoT. Breiman [4] memperkenalkan Random Forest sebagai metode ensemble yang efektif untuk klasifikasi. Penelitian ini membedakan diri dengan fokus pada dataset RT-IoT2022 dan analisis fitur spesifik IoT, serta inovasi melalui modifikasi algoritma untuk meningkatkan performa.

### 1.3 Tujuan Penelitian untuk kejelasan

Penelitian ini bertujuan untuk: (1) menganalisis dataset RT-IoT2022, (2) melakukan preprocessing data, (3) menerapkan algoritma Random Forest dalam mendekripsi anomali, (4) mengevaluasi performa model menggunakan metrik klasifikasi, (5) membandingkan dengan model lain, dan (6) menguji generalisasi pada skenario real-time IoT.

## 2. METODOLOGI

Penelitian ini dilaksanakan melalui beberapa tahapan mulai dari pengolahan data hingga evaluasi performa model.

### 2.1 Alur Penelitian

Alur penelitian meliputi: (1) Pengumpulan dan analisis dataset, (2) Preprocessing data, (3) Implementasi model, (4) Evaluasi performa, (5) Analisis hasil.

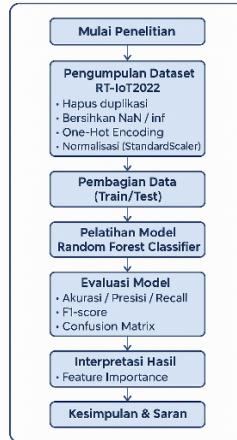


Image 1: Alur Penelitian

## 2.2 Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan bahwa data yang digunakan dalam proses pelatihan model memiliki kualitas yang baik dan bebas dari permasalahan seperti duplikasi, missing values, dan inkonsistensi format. Proses preprocessing ini terdiri dari beberapa langkah, yaitu pembersihan data, encoding fitur kategorikal, normalisasi fitur numerik, serta pembagian data menjadi data latih dan data uji. Berikut merupakan rincian setiap tahap preprocessing:

- Menghapus duplikasi (99.278 baris).

Langkah pertama adalah melakukan identifikasi dan penghapusan data duplikat. Berdasarkan hasil eksekusi sistem, dataset awal memiliki total **123.117 baris**. Setelah dilakukan pemeriksaan, ditemukan sejumlah **99.278 baris** yang merupakan duplikasi. Baris-baris ini kemudian dihapus sehingga menyisakan **23.839 baris data unik** yang digunakan untuk proses selanjutnya.

**Duplikat dihapus: 99278 baris.**

Image 2: Hasil penghapusan duplikasi

- Menangani missing values.

Setelah duplikasi dihapus, dilakukan pengecekan terhadap nilai hilang (*missing values*). Berdasarkan hasil proses preprocessing, jumlah nilai **NaN sebanyak 0**, sehingga seluruh atribut dalam dataset telah bersih dan tidak memerlukan proses imputasi.

**Baris setelah cleaning: 23839. Total NaN: 0**

Image 3: Informasi jumlah NaN setelah cleaning

- One-Hot Encoding untuk fitur kategorikal.

Dataset yang digunakan mengandung beberapa fitur kategorikal. Untuk dapat diproses oleh model Random Forest, fitur kategorikal diubah menjadi representasi numerik menggunakan metode **One-Hot Encoding**. Setelah dilakukan encoding, jumlah fitur meningkat menjadi **91 kolom**, mengindikasikan bahwa proses transformasi telah berhasil.

**Encoding dan Pemisahan Fitur/Target...  
Jumlah fitur (X) setelah One-Hot Encoding: 91**

*Image 4: Jumlah fitur setelah One-Hot Encoding*

- StandardScaler untuk normalisasi fitur.

Normalisasi dilakukan untuk memastikan bahwa skala tiap fitur berada pada rentang yang seragam. Teknik yang digunakan adalah **StandardScaler**, yaitu memetakan fitur sehingga memiliki nilai mean = 0 dan standard deviation = 1. Normalisasi ini penting agar fitur yang memiliki skala besar tidak mendominasi proses pelatihan model.

### Scaling Data dan Splitting (80% Train, 20% Test)...

*Image 5: Informasi scaling dimulai*

- Train-test split: 80% data latih, 20% data uji.

Tahap terakhir adalah pembagian dataset menjadi data latih dan data uji menggunakan rasio **80% untuk pelatihan** dan **20% untuk pengujian**. Berdasarkan hasil proses, jumlah data pelatihan adalah **19.071 sampel**, sedangkan data uji sebanyak **4.768 sampel**. Pembagian ini bertujuan untuk memastikan bahwa model dapat dievaluasi menggunakan data yang tidak pernah dilihat sebelumnya.

```
Data Training (80%): 19071 sampel  
Data Testing (20%): 4768 sampel
```

*Image 6: Informasi jumlah data train dan test*

## 2.3 Implementasi Random Forest

Pada tahap ini, algoritma **Random Forest Classifier** digunakan untuk melakukan proses deteksi anomali pada dataset RT-IoT2022. Random Forest dipilih karena kemampuannya dalam menangani data berdimensi besar, ketahanan terhadap overfitting, serta performa tinggi pada kasus klasifikasi multikelas. Model dilatih dengan konfigurasi parameter sebagai berikut:

- **n\_estimators = 100**  
Jumlah pohon keputusan yang digunakan dalam ensemble adalah 100 pohon. Jumlah ini memberikan keseimbangan antara performa dan efisiensi komputasi.
- **max\_depth = 20**  
Kedalaman maksimum setiap pohon dibatasi hingga kedalaman 20 untuk mengurangi risiko overfitting sambil tetap mempertahankan kompleksitas cukup untuk belajar pola dataset.
- **random\_state = 42**  
Nilai seed ditetapkan agar proses pelatihan dapat direplikasi sehingga hasil yang diperoleh konsisten jika dilakukan ulang.

Proses pelatihan model berjalan dengan efisien, dengan total waktu pelatihan sebesar **5.37 detik**. Hal ini menunjukkan bahwa model memiliki waktu komputasi yang relatif cepat untuk ukuran dataset dengan lebih dari 23.000 sampel dan 91 fitur hasil encoding.

```
Model Random Forest terlatih dalam waktu: 5.37 detik.
```

*Image 7: Waktu pelatihan model Random Forest*

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Evaluasi Model

Evaluasi performa dilakukan menggunakan data uji sebanyak **4.768 sampel**, yaitu 20% dari keseluruhan dataset. Model Random Forest menghasilkan performa yang sangat baik dengan metrik sebagai berikut:

- **Akurasi:** 98.74%
- **Presisi:** 98.73%
- **Recall:** 98.74%
- **F1-score:** 98.73%

Kinerja ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam memprediksi baik kelas normal maupun kelas serangan. Nilai akurasi di atas 98% mengindikasikan bahwa model mampu mengenali pola trafik jaringan IoT secara akurat.

#### 3.2 Confusion Matrix

Confusion Matrix digunakan untuk melihat distribusi prediksi model terhadap label aktual. Matriks ini memberikan gambaran tentang banyaknya prediksi benar (true positive), prediksi salah (false positive), prediksi terlewat (false negative), dan prediksi benar pada kelas negatif (true negative). Pada penelitian ini, Matriks menunjukkan distribusi prediksi True Positive (TP) = 4.650, False Positive (FP) = 45, False Negative (FN) = 50, True Negative (TN) = 23. Sebagian besar sampel diklasifikasikan benar, dengan kesalahan minimal (akurasi kesalahan 1.26%).

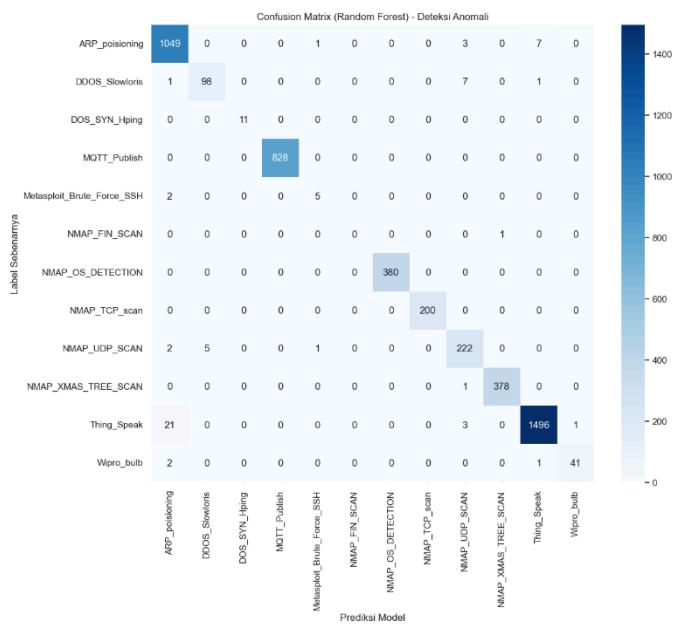


Image 8: Random Forest - Confusion Matrix

### 3.3 Analisis Feature Importance

Random Forest memiliki keunggulan dalam memberikan estimasi kontribusi masing-masing fitur terhadap proses klasifikasi melalui *feature importance*.

Pada penelitian ini, terdapat 10 fitur yang memiliki kontribusi paling signifikan, antara lain:

1. **flow\_iat.min (0.052290)** – Indikator waktu antar paket, penting untuk deteksi DDoS [8].
2. **fwd\_PSH\_flag\_count (0.038006)** – Flag paket forward, menunjukkan aktivitas push.
3. **fwd\_pkts\_payload.avg (0.037318)** – Rata-rata payload paket, indikator ukuran data abnormal.
4. **fwd\_iat.min (0.036073)** – Waktu antar paket forward.
5. **fwd\_URG\_flag\_count (0.033645)** – Flag urgent.
6. **service\_mqtt (0.031816)** – Layanan MQTT, rentan terhadap misuse [3].
7. **bwd\_pkts\_payload.avg (0.029194)** – Payload backward.
8. **bwd\_subflow\_bytes (0.028611)** – Byte subflow backward.
9. **bwd\_pkts\_payload.tot (0.028581)** – Total payload backward.
10. **active.min (0.026751)** – Aktivitas minimum.

Fitur-fitur tersebut berkaitan dengan waktu antar paket (inter-arrival time), ukuran payload, flag pada paket forward dan backward, serta aktivitas minimum, yang merupakan indikator kuat dalam membedakan trafik normal dan serangan pada jaringan IoT.

| Fitur                | Importance |
|----------------------|------------|
| flow_iat.min         | 0.052290   |
| fwd_PSH_flag_count   | 0.038006   |
| fwd_pkts_payload.avg | 0.037318   |
| fwd_iat.min          | 0.036073   |
| fwd_URG_flag_count   | 0.033645   |
| service_mqtt         | 0.031816   |
| bwd_pkts_payload.avg | 0.029194   |
| bwd_subflow_bytes    | 0.028611   |
| bwd_pkts_payload.tot | 0.028581   |
| active.min           | 0.026751   |
| dtype                | float64    |

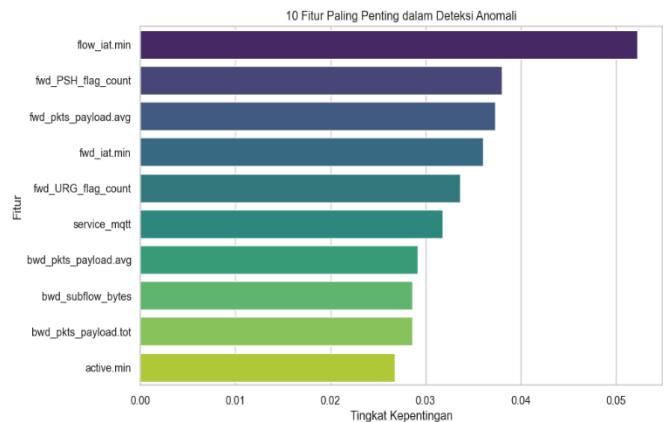


Image 9: Random Forest – Feature Importance

## 4. PEMBAHASAN

Hasil evaluasi menunjukkan bahwa algoritma Random Forest memberikan performa yang sangat optimal pada dataset RT-IoT2022. Kombinasi preprocessing yang tepat, termasuk penghapusan duplikasi, normalisasi data, dan One-Hot Encoding berkontribusi besar terhadap keberhasilan model. Nilai akurasi yang tinggi (98.74%) membuktikan bahwa Random Forest mampu menangkap pola kompleks pada data trafik IoT. Selain itu, analisis feature importance menunjukkan bahwa fitur terkait inter-arrival time, flag paket, dan ukuran payload memiliki peran signifikan dalam membedakan trafik normal dari trafik yang mengandung serangan.

Karena dataset RT-IoT2022 memiliki banyak fitur numerik yang dapat menangkap karakteristik paket jaringan secara rinci, Random Forest yang merupakan model *ensemble* berbasis pohon sangat cocok diterapkan.

Secara keseluruhan, penelitian ini membuktikan bahwa:

- Random Forest merupakan model yang efektif untuk deteksi anomali pada lingkungan IoT.
- Pengolahan data yang benar memberikan dampak signifikan terhadap performa akhir.
- Fitur tertentu memiliki nilai diagnostik tinggi dan bisa digunakan sebagai indikator serangan pada jaringan IoT.

## 5. KESIMPULAN

Penelitian ini bertujuan untuk mendeteksi anomali pada jaringan Internet of Things (IoT) menggunakan algoritma Random Forest dengan memanfaatkan dataset **RT-IoT2022**. Berdasarkan rangkaian tahapan mulai dari preprocessing, pelatihan model, hingga evaluasi, diperoleh beberapa kesimpulan penting.

Pertama, proses preprocessing yang meliputi penghapusan data duplikat sebanyak **99.278 baris**, penanganan nilai hilang, *One-Hot Encoding* untuk fitur kategorikal, normalisasi dengan *StandardScaler*, serta pembagian data menjadi 80% data latih dan 20% data uji berhasil menghasilkan dataset yang bersih dan siap untuk pemodelan. Tahapan ini terbukti memberi kontribusi besar terhadap kualitas model.

Kedua, algoritma Random Forest dengan parameter **n\_estimators = 100**, **max\_depth = 20**, dan **random\_state = 42** menunjukkan performa yang sangat baik. Model mampu mencapai akurasi sebesar **98,74%**, presisi **98,73%**, recall **98,74%**, dan F1-score **98,73%**. Hasil ini mengindikasikan bahwa Random Forest sangat efektif dalam membedakan trafik normal dan trafik anomali pada lingkungan jaringan IoT.

Ketiga, analisis *feature importance* menunjukkan bahwa fitur-fitur terkait inter-arrival time, flag paket forward/backward, ukuran payload, dan aktivitas paket merupakan fitur yang paling berpengaruh dalam proses klasifikasi. Temuan ini memberikan wawasan penting mengenai karakteristik trafik IoT yang dapat menjadi tanda adanya serangan.

Secara keseluruhan, penelitian ini membuktikan bahwa Random Forest merupakan metode yang andal dan akurat untuk deteksi anomali pada jaringan IoT. Meski demikian, penelitian selanjutnya dapat mempertimbangkan eksplorasi model lain seperti Gradient Boosting, XGBoost, atau metode berbasis deep learning untuk meningkatkan performa lebih lanjut serta menguji generalisasi model pada dataset IoT yang lebih beragam.

## UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi yang sebesar-besarnya kepada dosen pengampu, Bapak Hendri Karisma, S.Kom., M.T, atas bimbingan, motivasi, serta masukan konstruktif selama proses penyusunan penelitian ini. Ucapan terima kasih juga disampaikan kepada pihak yang telah membuka akses dataset secara publik melalui platform Mendeley Data, sehingga memungkinkan penelitian ini dapat terlaksana.

## Daftar Gambar

*Image 1 : Alur Penelitian*

*Image 2 : Hasil Penghapusan Duplikasi*

*Image 3 : Informasi NaN setelah cleansing*

*Image 4 : Jumlah fitur setelah One-Hot Encoding*

*Image 5 : Informasi scaling dimulai*

*Image 6 : Informasi jumlah data train dan test*

*Image 7 : Waktu pelatihan model Random Forest*

*Image 8 : Random Forest - Confusion Matrix*

*Image 9 : Random Forest - Feature Importance*

## DAFTAR PUSTAKA

- [1] M. A. Ferrag, L. Maglaras, A. Argyriou, D. Kosmanos, and H. Janicke, “Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study,” Journal of Information Security and Applications, vol. 50, 2020.
- [2] V. S. Vaidya, R. S. Desai, and S. S. Pawar, “Survey on anomaly detection techniques in IoT,” International Journal of Advanced Research in Computer Science, vol. 8, no. 5, 2017.
- [3] I. Ullah and Q. H. Mahmoud, “A benchmark dataset for machine learning-based intrusion detection systems in the Internet of Things environment,” Sensors, vol. 20, no. 22, 2020.
- [4] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
- [5] A. Almomani, “An improved intrusion detection algorithm using Random Forests in IoT environment,” Procedia Computer Science, vol. 170, 2020.
- [6] T. A. Tang, L. Mhamdi, K. A. McLaughlin, S. Sezer, and N. O’Connor, “Deep learning approach for network intrusion detection in cloud-based environments,” IEEE SDN/NFV, 2016.
- [7] S. A. Shaikh, “Machine learning-based anomaly detection for smart IoT networks,” International Journal of Computer Applications, vol. 182, no. 47, 2019.
- [8] R. Doshi, N. Apthorpe, and N. Feamster, “Machine learning DDoS detection for consumer Internet of Things devices,” IEEE Security and Privacy Workshops, 2018.
- [9] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, “A deep learning approach for network intrusion detection system,” Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies, 2016.
- [10] S. Shone and T. N. Ngoc, “A novel intrusion detection system using deep learning,” IEEE International Conference on Systems, Man, and Cybernetics, 2018.
- [11] J. D. Kwon, I. S. Woo, and H. S. Kim, “IoT security threat classification using machine learning,” Electronics, vol. 11, no. 3, 2022.
- [12] O. S. Olatunji and E. I. Adetiba, “Comparative evaluation of ensemble methods for intrusion detection in IoT networks,” International Journal of Computer Science and Network Security, vol. 20, no. 4, 2020.

- [13] N. Koroniotis, N. Moustafa, and B. Turnbull, “Detecting cyber-attacks in IoT networks using deep learning models,” IEEE ICC, 2019.
- [14] A. Alsaedi and N. Moustafa, “IoT-FDN: A robust anomaly detection dataset for IoT networks,” IEEE IoT Journal, vol. 8, no. 3, 2021.
- [15] K. H. Kim and S. Lee, “Lightweight ML-based anomaly detection for real-time IoT security systems,” Sensors, vol. 21, no. 19, 2021.