PROPOSAL

Anomaly Detection in Internet of Things (IoT) Network with Rt-IoT2022 Dataset using Random Forest

Dosen Pengampu : Hendri Kharisma S.Kom, M.T

Disusun Oleh :

Aisyah (241572010003)

Prameswari Kirana Jingga (241572010021)

Sekolah Tinggi Manajemen Informatika dan Komputer TAZKIA.

Jl. Raya Dramaga Blok Radar Baru No.8, RT.03/RW.03, Margajaya, Kec. Bogor Barat, Kota Bogor, Jawa Barat 16116, Indonesia

ABSTRAK

Perkembangan Internet of Things (IoT) membawa kemajuan signifikan dalam konektivitas dan otomatisasi di berbagai sektor. Namun, peningkatan konektivitas ini juga memunculkan ancaman keamanan jaringan yang semakin kompleks. Penelitian ini bertujuan untuk mendeteksi anomali dalam jaringan IoT menggunakan algoritma Random Forest berbasis supervised learning. Dataset yang digunakan adalah RT-IoT2022 dari UCI Machine Learning Repository, yang berisi data trafik jaringan dari berbagai perangkat IoT dengan label normal dan attack. Tahapan penelitian meliputi pengumpulan data, preprocessing, pelatihan model Random Forest, serta evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan sistem Intrusion Detection System (IDS) yang efisien, akurat, dan mudah diimplementasikan pada lingkungan IoT.

Kata kunci: Internet of Things, Keamanan Jaringan, Random Forest, Deteksi Anomali, RT-IoT2022.

BAB I PENDAHULUAN

1.1 Latar Belakang

Internet of Things (IoT) merupakan jaringan yang menghubungkan berbagai perangkat fisik melalui internet sehingga memungkinkan pertukaran data secara otomatis. IoT digunakan di berbagai bidang seperti industri, kesehatan, transportasi, dan rumah tangga pintar. Namun, konektivitas tinggi ini juga menimbulkan risiko terhadap keamanan data dan privasi pengguna. Banyak perangkat IoT memiliki sumber daya terbatas sehingga tidak mampu menjalankan sistem keamanan kompleks, sehingga mudah menjadi target serangan siber seperti DDoS, Port Scanning, dan Brute-Force Attack.

Untuk mengatasi hal tersebut, dibutuhkan sistem deteksi anomali yang mampu mengenali pola serangan dengan akurasi tinggi. Algoritma Random Forest dipilih karena memiliki keunggulan dalam menangani data berlabel besar, mengurangi overfitting, dan memberikan interpretasi terhadap pentingnya fitur. Dataset RT-IoT2022 digunakan karena merupakan kumpulan data terbaru yang menggambarkan trafik nyata jaringan IoT dari berbagai perangkat.

1.2 Rumusan Masalah

1. Bagaimana karakteristik data pada dataset RT-IoT2022?
2. Bagaimana penerapan algoritma Random Forest untuk deteksi anomali pada jaringan IoT?
3. Seberapa akurat model Random Forest dalam mengklasifikasikan trafik normal dan serangan?

1.3 Tujuan Penelitian

1. Menganalisis karakteristik dataset RT-IoT2022.
2. Mengimplementasikan algoritma Random Forest untuk deteksi anomali jaringan IoT.
3. Mengevaluasi performa model menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score.

1.4 Manfaat Penelitian

1. Menghasilkan model deteksi anomali berbasis supervised learning yang akurat dan efisien.
2. Memberikan solusi praktis terhadap tantangan keamanan pada lingkungan IoT.
3. Menjadi referensi untuk penelitian lanjutan dalam pengembangan sistem keamanan berbasis data.

BAB II DESKRIPSI DATASET

Dataset yang digunakan dalam penelitian ini adalah RT-IoT2022, yang bersumber dari UCI Machine Learning Repository https://archive.ics.uci.edu/dataset/942/rt-iot2022?utm_source=chatgpt.com

Dataset ini dikembangkan oleh Sharmila dan Nagapadma (2023) dan berisi data trafik jaringan dari berbagai perangkat IoT nyata seperti sensor suhu, lampu pintar, dan asisten suara. Dataset mencakup dua kategori utama, yaitu trafik normal dan serangan (attack), dengan beberapa jenis serangan seperti DDoS, Port Scanning, dan SSH Brute Force.

Jumlah total record pada dataset mencapai lebih dari 600.000 entri dengan 24 hingga 70 fitur yang menggambarkan karakteristik aliran data (flow-based features). Setiap fitur merepresentasikan parameter jaringan seperti durasi koneksi, jumlah paket masuk dan keluar, serta rata-rata byte per detik.

BAB III METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Penelitian ini akan dilakukan melalui lima tahapan utama:

1. **Pengumpulan Data:**
   Mengambil dataset dari **UCI Machine Learning Repository**. Dataset yang digunakan adalah **RT-IoT2022**, yang berisi data trafik jaringan dari berbagai perangkat IoT dengan label *normal* dan *attack* (serangan). Dataset ini dipilih karena menggambarkan kondisi nyata jaringan IoT dan relevan untuk penelitian deteksi anomali.

2. **Preprocessing Data:**
   Dilakukan untuk mempersiapkan dataset sebelum masuk ke tahap pelatihan model. Langkah-langkah preprocessing meliputi:

   - **Penanganan missing values** dengan metode imputasi (mengisi nilai kosong) atau penghapusan baris/kolom yang tidak relevan.
   - **Konversi data kategorikal menjadi numerik** menggunakan teknik *label encoding* atau *one-hot encoding*.
   - **Normalisasi nilai numerik** agar skala tiap fitur seimbang.
   - **Split dataset** menjadi dua bagian, yaitu 80% untuk *training* dan 20% untuk *testing*.

3. **Pemodelan (Modeling):**
   Pada tahap ini dilakukan pembangunan model deteksi anomali menggunakan algoritma

Random Forest.

- **Algoritma:** Random Forest (Supervised Learning).
- **Tools:** Python (pandas, scikit-learn, matplotlib, seaborn).
- **Parameter utama:** jumlah pohon (*n_estimators*) 100, kedalaman maksimum (*max_depth*) 15–25, dan kriteria pemisahan *gini impurity*.
  Model dilatih menggunakan data *training* untuk mengenali pola trafik normal dan serangan, kemudian diuji menggunakan data *testing*.

4. **Evaluasi Model:**
   Evaluasi dilakukan untuk mengukur performa model dalam mendeteksi anomali jaringan IoT.
   - **Metrik evaluasi:** Accuracy, Precision, Recall, F1-score, dan Confusion Matrix.
   - Selain itu, dilakukan analisis **feature importance** untuk mengetahui fitur yang paling berpengaruh terhadap hasil klasifikasi.
5. **Visualisasi & Analisis Hasil:**
   Setelah evaluasi, hasil model divisualisasikan dan dianalisis untuk mendapatkan interpretasi yang lebih baik.
   - Pembuatan *heatmap* korelasi fitur untuk melihat hubungan antar variabel.
   - Visualisasi Confusion Matrix dan diagram perbandingan nilai metrik model.
   - Analisis hasil dilakukan untuk menilai efektivitas model dan potensi pengembangannya pada penelitian berikutnya.

3.2 Alur Penelitian

Alur kerja penelitian ini dapat digambarkan sebagai berikut:

1. Input Dataset (RT-IoT2022 dari UCI Repository)
2. Preprocessing (pembersihan, encoding, normalisasi, pembagian data)
3. Training Model (pelatihan menggunakan Random Forest)
4. Evaluation (pengujian model dan analisis metrik)
5. Visualization & Interpretation (analisis visual dan interpretasi hasil)
6. Conclusion (kesimpulan dan rekomendasi penelitian)

BAB V TINJAUAN PUSTAKA

1. Sharmila, B. S., & Nagapadma, R. (2023). Quantized Autoencoder (QAE) Intrusion Detection System for Anomaly Detection in Resource-Constrained IoT Devices Using RT-IoT2022 Dataset. Cybersecurity, 6(41). SpringerOpen. DOI: 10.1186/s42400-023-00178-5.
2. UCI Machine Learning Repository. (2024). RT-IoT2022: Real-Time Internet of Things Dataset. University of California, Irvine.
3. Breiman, L. (2001). Random Forests. Machine Learning Journal, 45(1), 5–32.
4. Ring, M. et al. (2019). A Survey of Network-Based Intrusion Detection Datasets. Computers & Security, 86, 147–167.