Attack Detection On Edge IIot Using Ensemble Learning And XAI Model Transparency

Hussain Randika[1], Parman Sukarno[2], Aulia Arif Wardana[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[4]Divisi Digital Service PT Telekomunikasi Indonesia
[1]hrandika@students.telkomuniversity.ac.id,
[2]Psukarno@telkomuniversity.ac.id,[3]auliawardan@telkomuniversity.ac.id,

Abstrak

Industrial Internet of Things (IIoT) telah berkembang pesat, meningkatkan efisiensi produksi dan profitabilitas, sekaligus memperkenalkan tantangan keamanan yang signifikan karena sistem IIoT menjadi lebih rentan terhadap serangan siber. Studi ini menangani tantangan tersebut dengan menerapkan deteksi serangan menggunakan metode ensemble learning dan Explainable Artificial Intelligence (XAI). Model seperti Random Forest, XGBoost, dan LightGBM digunakan bersama SHAP (SHapley Additive Explanations) untuk meningkatkan transparansi model.

Stratified Sampling digunakan untuk mengurangi volume data sambil mempertahankan distribusi fitur, dan seleksi fitur dilakukan menggunakan Random Forest Feature Importance untuk mencapai akurasi dan efisiensi tinggi. Dataset Edge-IIoTset digunakan, dengan preprocessing dan seleksi fitur yang dioptimalkan melalui Random Forest Feature Importance. Pengujian melibatkan dua skenario dengan variasi jumlah fitur dan pembagian data untuk mengevaluasi efektivitas model.

Hasil menunjukkan bahwa model LightGBM mencapai akurasi tertinggi sebesar 97,60%, diikuti oleh XGBoost dengan 96,90%, dan Random Forest dengan 96,51%. Selain itu, SHAP mengidentifikasi fitur-fitur utama yang memengaruhi prediksi, meningkatkan kepercayaan dan pemahaman pengguna. Hasil ini menunjukkan bagaimana kombinasi ensemble learning dan XAI dapat meningkatkan keamanan IIoT secara efektif.

Kata kunci : Explainable AI, Industrial Internet of Things, Feature Selection, Stratified Sampling, Ensemble Learning

Abstract

The Industrial Internet of Things (IIoT) has rapidly developed, improving production efficiency and profitability and introducing significant security challenges as IIoT systems become more vulnerable to cyberattacks. This study addresses these challenges by implementing attack detection using en semble learning methods and Explainable Artificial Intelligence (XAI). Models like Random Forest, XGBoost and LightGBM are employed with SHAP (SHapley Additive Explanations) to improve model transparency. Stratified Sampling reduces data volume while preserving feature distribution, and feature selec tion is performed using Random Forest Feature Importance to achieve high accuracy and efficiency. The Edge-IIoTset dataset is used, with preprocessing and feature selection optimized through Random Forest Feature Importance. Testing involves two sce narios with varying feature counts, and data splits to evaluate effectiveness. Results show that the LightGBM model achieved the highest accuracy 97.60%, followed by XGBoost 96.90% and Random Forest 96.51%. Additionally, SHAP identified key features influencing predictions and improving user trust and understanding. These results demonstrate how well ensemble learning and XAI work together to improve IIoT security.

Keywords : Explainable AI, Industrial Internet of Things, Feature Selection, Stratified Sampling, Ensemble Learning

1. Pendahuluan

1.1 Latar Belakang

IIoT, atau Industrial Internet of Things, telah menjadi bagian penting dalam perkembangan era Industri 4.0, memungkinkan aktivitas industri menjadi lebih otomatis dan saling terhubung. Sebagai salah satu teknologi pendukung IoT, Edge Computing dalam IIoT merupakan konsep yang muncul dari kemajuan pesat IoT [1]. Edge Computing memungkinkan pemrosesan data lebih dekat ke sumbernya, sehingga mengurangi latensi dan

meningkatkan efisiensi operasional. Namun, selain manfaat tersebut, kemajuan dalam IoT juga menghadirkan tantangan keamanan yang signifikan. Serangan terhadap IIoT dapat mengganggu transfer data yang dapat dipercaya dan efektif antara proses industri yang menggunakan kecerdasan buatan dalam sistem siber-fisik. Ekosistem IIoT memungkinkan peralatan aplikasi industri berkomunikasi dengan sedikit bantuan dari manusia [2]. Serangan siber terhadap perangkat IoT telah meningkat secara signifikan, dengan banyak pelanggaran yang tercatat dalam waktu singkat [3].

Saat ini, banyak solusi deteksi serangan untuk IoT dan IIoT masih bergantung pada metode tradisional, seperti deteksi berbasis tanda tangan (*signature-based detection*) dan analisis statis, yang sering kali tidak memadai dalam menghadapi kompleksitas serangan siber yang semakin berkembang. Model kecerdasan buatan yang diterapkan dalam deteksi serangan sering kali berbentuk "kotak hitam" (*black box*), sehingga sulit bagi pengguna untuk memahami dan mempercayai keputusan yang dibuat. Tantangan ini menyoroti perlunya pendekatan yang lebih canggih dan transparan dalam deteksi serangan. *Ensemble learning* telah muncul sebagai pendekatan yang kuat untuk meningkatkan efektivitas sistem deteksi intrusi dalam lingkungan edge IIoT. *Ensemble learning* melibatkan penggunaan beberapa algoritma pembelajaran untuk meningkatkan kinerja prediksi dibandingkan dengan menggunakan satu algoritma pembelajaran saja [4][5]. Tujuan utama penggunaan *ensemble learning* dalam pelatihan model adalah meningkatkan akurasi deteksi ancaman dalam sistem siber-fisik dengan mengatasi ketidakseimbangan kelas dan meningkatkan kinerja prediksi secara keseluruhan. *Ensemble learning* berusaha meningkatkan akurasi dibandingkan metode pembelajaran mesin biasa dengan mengombinasikan berbagai algoritma [6].

Dalam penelitian ini, kami menerapkan pendekatan *ensemble learning* menggunakan model Random Forest, XGBoost, dan LightGBM untuk menentukan algoritma yang paling efektif dalam mendeteksi serangan siber terhadap IIoT. Model-model ini digabungkan untuk memaksimalkan kemampuan masing-masing, dengan tujuan memberikan solusi yang lebih akurat dan tangguh dalam mendeteksi serangan terhadap perangkat IIoT. Selain *ensemble learning*, *Explainable Artificial Intelligence* (XAI) semakin penting dalam meningkatkan transparansi model deteksi [6]. Dengan menggunakan teknik SHAP (*SHapley Additive ExPlanations*), kami berharap dapat memberikan penjelasan yang jelas dan mudah dipahami tentang bagaimana model membuat keputusan berdasarkan hasil akhir. Transparansi ini memungkinkan profesional keamanan untuk lebih memahami alasan di balik prediksi model, menilai keadilan dan keandalannya, serta mengidentifikasi area yang perlu ditingkatkan dalam sistem deteksi. Penelitian ini bertujuan untuk mengembangkan alat deteksi serangan yang kuat untuk sistem IIoT dengan menggunakan *ensemble learning* yang didukung oleh wawasan dari XAI. Pendekatan ini dikembangkan untuk meningkatkan akurasi deteksi serangan sekaligus membuat proses pengambilan keputusan terhadap hasil akhir menjadi lebih transparan dan mudah dipahami oleh pengguna. Dataset Edge-IIoTset digunakan untuk melatih dan memvalidasi model, dengan menyediakan berbagai contoh serangan IIoT guna membantu merancang sistem deteksi yang kuat.

Penelitian ini dilakukan untuk menutup kesenjangan utama dalam sistem deteksi ancaman IIoT yang ada, yang sering kali kurang transparan dan tidak efektif dalam menjelaskan hasil model. Dengan menggabungkan ensemble learning dan XAI, penelitian ini menghadirkan strategi yang lebih dapat diinterpretasikan dan lebih akurat dalam mengidentifikasi ancaman di lingkungan edge IIoT.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan sebelumnya, maka rumusan masalah yang akan dibahas adalah sebagai berikut:

1. Bagaimana hasil penerapan *Ensemble Learning* dalam mendeteksi serang-an pada *Industrial Internet Of Things*?
2. Bagaimana XAI dapat menjelaskan hasil akurasi dari metode yang digunakan?

1.3 Tujuan

Adapun tujuan dari penelitian ini, antara lain :

1. Menerapkan *Ensemble Learning* dalam mendeteksi serangan pada Industrial Internet Of Things.
2. Menerapkan XAI untuk menjelaskan hasil akurasi dari pelatihan Algoritma yang telah digunakan.