Media Journal of General Computer Science, **2023, 1, 14-19**
Published Online January 2024 in MASE (http://www.mase.or.id)
DOI: 10.62205/mjgcs.v1i1.8

Improvement Attack Detection on Internet of Thinks Using Principal Component Analysis and Random Forest

Adrian Pirtama
Department of Informatics, Dinamika Bangsa University, Jambi, Indonesia
E-mail: adrianpirtama2020@gmail.com

Yuda Prasetia
Department of Informatics, Dinamika Bangsa University, Jambi, Indonesia
E-mail: yudaprass18@gmail.com

Redho Irnindo Saputra
Department of Informatics, Dinamika Bangsa University, Jambi, Indonesia
E-mail: redhoirnindo@gmail.com

Eko Arip Winanto
Department of Computer Engineering, Dinamika Bangsa University, Jambi, Indonesia
E-mail: ekoaripwinanto@gmail.com
*Corresponding Author

Received: 13 November, 2023; Accepted: 23 November, 2023; Published: 30 January, 2024

Abstract: Network security has become crucial in facing increasingly complex and sophisticated attack threats. Network intrusion detection aids in identifying suspicious activities indicating unauthorized intrusions. This research aims to enhance the performance of advanced attack detection. The Random Forest method is an algorithm that leverages an ensemble of decision trees. This ensemble comprises several independent decision trees used to classify data. One characteristic of the Random Forest method is its ability to address overfitting issues and provide good predictive quality. One approach to improving RF's performance is through Principal Component Analysis (PCA). PCA is a statistical technique used to reduce feature dimensionality. PCA eliminates feature correlations and identifies essential features that can enhance the detection of attacks and normal traffic. This research will be tested with the CIC IoT 2023 dataset, encompassing various attack types. The model testing consists of four feature dimensions, namely 5, 8, 10, and 47. The detection results are promising, significantly improving attack detection performance, reaching up to 99.2%.

Keywords: IoT, IDS, PCA, Random Forest

I. Introduction

Peningkatan performa deteksi serangan merupakan hal yang sangat penting dalam keamanan komputer dan jaringan. Dalam era digital yang semakin maju, serangan terhadap sistem komputer dan jaringan semakin kompleks. Beberapa kasus terjadi pada Badan usaha milik pemerintah (BUMN), salah satunya yang baru saja terjadi pada Bank Syariah Indonesia (BSI). Bank Syariah Indonesia terus melakukan upaya perbaikan setelah beberapa hari layanan ATM, *mobile* dan *online banking error* dan layanan tidak dapat diakses. Oleh sebab itu, penting untuk mendeteksi serangan pada jaringan, sehingga serangan dapat dideteksi dan dilawan sebelum dapat merusak sistem. Salah satu cara untuk mengatasi masalah ini adalah dengan mengembangkan sistem deteksi intrusi (*intrusion detection systems*/IDS), *firewall*, dan alat-alat lainnya. Pada penelitian ini akan fokus pada penerapan IDS untuk mengidentifikasi komunikasi data yang mencurigakan atau tidak normal. [1]

Masalah yang ada pada sistem deteksi adalah bagaimana meningkatkan kinerja dari sistem deteksi. Beberapa penelitian sebelumnya [2] [3] telah mengusulkan penelitian untuk meningkatkan kinerja dari sistem deteksi serangan di jaringan. Penelitian [2] menggusulkan optimasi algoritma *random forest* menggunakan *principal component* analysis untuk deteksi malware. Penelitian ini membahas tentang peningkatan performa algoritma Random Forest menggunakan PCA. *Random Forest* dipilih karena memiliki performa Akurasi dari *Recall* terbaik dibandingkan 4 algoritma lain, seperti: *Adaboost, Neural Network, Support Vector Machine* dan *kNearest Neighbor*. Performa akurasi yang diperoleh *random forest* pada *dataset malware* adalah 98.3% maupun *Recall*. Selanjutnya eksperimen dilakukan dengan melakukan fitur reduksi pada *dataset*, hasil dari fitur direduksi menjadi 32 maka performa algoritma *random forest* meningkat menjadi 98.44%. Berdasarkan hasil dari pengurangan fitur menggunakan PCA berhasil meningkat performa akurasi. Hal ini menunjukkan adanya peningkatan pada algoritma *Random Forest*, karena semakin sedikit jumlah kesalahan deteksi.

Lalu penelitian selanjutnya [3] menerapan metode *principal component analysis* (PCA) untuk deteksi anomali pada jaringan *peer-to-peer* (P2P) *botnet.* Evaluasi dan hasil yang diperoleh dalam penelitian ini menunjukkan performa yang memuaskan dalam mendeteksi adanya anomali pada jaringan. Kemudian pada penelitian [4] telah menggusulkan deteksi serangan *lrddos* pada *sd-iot* menggunakan *random forest* dengan logistik koefisien regresi. Kekurangan system deteksi pada jaringan *IoT* tingkat rendah adalah masalah yang paling signifikan untuk diselesikan yaitu dalam hal manajemen terpusat. Sehingga memungkinkan terjadinya celah yang dapat dimanfaatkan oleh penyerang untuk mengambil alih kontrol secara keseluruhan. Eksperimen ini mengusulkan seleksi fitur menggunakan *Logistic Regression Koefisien* dan metode *random forest* yang digunakan untuk prediksi *DDoS* tingkat rendah memiliki akurasi tertinggi sebesar 98,7%. Oleh karena itu, pada penelitian ini menggusulkan metode *random forest* untuk mendeteksi serangan pada jaringan IoT dengan hibrid metode PCA untuk mengurangi jumlah fitur dataset.

Random forest merupakan salah satu metode klasifikasi yang sering digunakan dalam berbagai penelitian dan kasus pemodelan. Metode ini melibatkan pembuatan keputusan berdasarkan pembentukan pohon keputusan atau *decision tree*. Setiap cabang dalam pohon tersebut mengandung pertanyaan yang digunakan untuk memecahkan suatu keputusan berdasarkan jumlah cabang yang ideal [4]. Sehingga memiliki potensi untuk digunakan sebagai metode deteksi serangan. Pada penelitian [5] mengusulkan penerapan metode *feature extraction* untuk proses pemilihan fitur. Tujuanya adalah untuk menggurangi dimensi dari fitur dataset. Sehingga dapat meningkatkan akurasi dari sistem deteksi serangan. Salah satu metode *feature extraction* adalah PCA digunakan untuk mencari fitur yang mempengaruhi serangan pada jaringan. PCA adalah teknik yang handal untuk mengekstraksi struktur dari suatu dataset [6]. Oleh karena itu pada penelitian ini menggusulkan metode *hybrid Random Forest Classifier* untuk metode deteksi dan metode PCA sebagai ekstraksor factor yang digunakan untuk menghasilkan deteksi yang ideal [7][8].

Sistematika penulisan pada penelitan ini selanjutnya adalah pada bagian kedua berisi metode yang disusulkan. Ketiga adalah hasil dan diskusi, serta terakhir adalah kesimpulan.

2. Research Method

2.1. Experiment Setup

Tujuan utama dari penelitian ini adalah untuk mengidentifikasi dan mendeteksi serangan yang mungkin terjadi dalam jaringan. Untuk mencapai tujuan tersebut, penelitian ini menerapkan metode Random Forest sebagai alat utama dalam proses deteksi serangan. Sejumlah tahap penelitian yang perlu dilakukan akan diuraikan secara lebih rinci dalam alur penelitian yang ditampilkan dalam Fig.1.

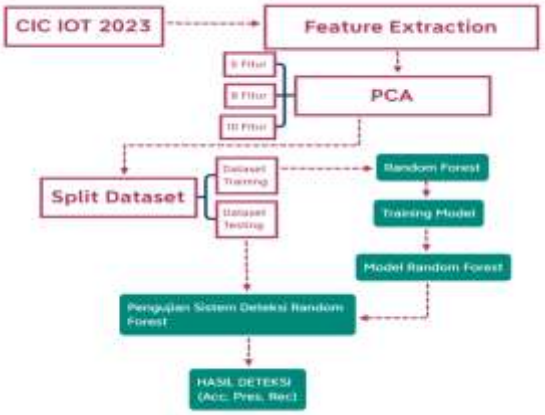

Fig.1. Experiment setup

Fig.1 adalah diagram eksperimen yang dirancang khusus untuk penelitian ini. Oleh karena itu, proses penelitian dibagi menjadi tiga tahap terpisah.

- Sebagai langkah awal, dilakukan ekstraksi fitur dari kumpulan data menggunakan metode PCA dan kumpulan data tersebut dibagi menjadi kumpulan data pelatihan dan kumpulan data pengujian.
- Pada tahap kedua, dilakukan pelatihan model menggunakan kumpulan data pelatihan untuk mendapatkan hasil model Random Forest untuk mendeteksi serangan cyber.
- Pada langkah terakhir, dilakukan pengujian menggunakan dataset pengujian pada model yang dibuat pada langkah sebelumnya untuk menghitung tingkat keberhasilan seperti akurasi, presisi, dan recall.

2.2. Dataset

Penelitian ini akan memanfaatkan dataset CIC IoT 2023 sebagai sumber data. Dataset tersebut mengklasifikasikan serangan menjadi tujuh kategori yang berbeda, yakni DDoS, DoS, Recon, Web-based, Brute Force, Spoofing, dan Mirai. Seluruh serangan yang tercatat dalam dataset ini dilakukan oleh perangkat IoT berbahaya yang memiliki tujuan menyerang perangkat IoT lainnya [9].

2.3. Proposed Method

Ekstraksi fitur memiliki peran penting dalam mengidentifikasi aspek-aspek kunci dari data, menghasilkan keefektifan, kemudahan, dan kejelasan. Proses ekstraksi fitur memengaruhi signifikan kualitas klasifikasi, terutama dalam Sistem Deteksi Intrusi (IDS), yang responsnya berfluktuasi sesuai dengan fitur-fitur yang digunakan sebagai input. Selain itu, hasil klasifikasi juga dipengaruhi oleh lingkungan lalu lintas yang padat, seperti yang sering terjadi dalam jaringan kompleks seperti Internet of Things, serta penggunaan fitur multidimensi. Oleh karena itu, penelitian ini menerapkan metode PCA untuk mereduksi dimensi dataset. Pseudocode PCA yang digunakan dalam penelitian dijelaskan secara rinci di bawah ini. Dataset dengan berbagai dimensi fitur, yaitu 5, 8, 10, dan 47, akan digunakan dalam proses pelatihan dan deteksi. Pada Fig.2 adalah proses dari random forest untuk deteksi sarangan pada penelitian ini.

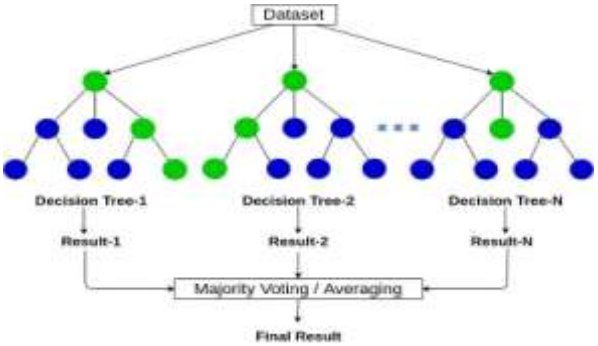

Fig.2. Arsitektur Random Forest [10]

Algoritma Random Forest memiliki kapasitas untuk melakukan klasifikasi pada data yang memiliki atribut yang tidak lengkap, dan sangat sesuai digunakan untuk mengklasifikasikan data sampel yang besar. Dalam proses klasifikasi Random Forest, data sampel akan secara acak dibagi ke dalam pohon keputusan. Setiap pohon yang terbentuk memiliki node akar, node internal, dan node leaf [11].

2.4. Environment Setup

Eksperimen dalam penelitian ini dilaksanakan pada sebuah laptop dengan spesifikasi tertentu, yaitu menggunakan sistem operasi Windows 10 Pro 64-bit (versi 10.0, Build 19045), dilengkapi dengan prosesor AMD Ryzen 5 PRO 2500U yang memiliki Radeon Vega Mobile Gfx (dengan 8 CPU) dan kecepatan sekitar 2.0 GHz, serta RAM berkapasitas 16 GB. Selain itu, alat-alat analisis yang diperlukan dalam penelitian meliputi Python, Scikit-Learn, TensorFlow, dan Keras.

3. Results and Discussion

Bagian ini mencakup hasil eksperimen yang telah dilaksanakan, termasuk laporan hasil dari langkah-langkah reduksi fitur dan uji kinerja algoritma Random Forest. Dalam pembahasan ini, akan dievaluasi efektivitas proses reduksi fitur serta analisis kinerja algoritma Random Forest.

3.1. Hasil Pemilihan Fitur

Pada proses ekstraksi fitur ini, algoritma *Random Forest* akan diperbaiki dengan menerapkan metode *Principal Component Analysis* (PCA) untuk mengurangi jumlah fitur. Tujuannya adalah untuk mengurangi upaya komputasi dan meningkatkan kinerja sistem pengenalan di jaringan IoT tanpa kehilangan karakteristik data. Tabel 2 menunjukkan hasil

ekstraksi ciri yang dilakukan dengan metode PCA. Dari tabel hasil, kita dapat melihat 3 atribut dikonversi dalam pencarian ini. Hasil reduksi fitur ini digunakan untuk mengolah data latih IDS menggunakan algoritma Random Forest.

Tabel 2. Hasil pemilihan fitur

Jumlah fitur	Hasil PCA
5	191565,-7, 9731.7, 24450.34, -13594, 546.36132
8	191565 -7, 9731.7,24450.34, -13594, 546.3613, -5.11122,-0.13944, 8.075055
10	191565 -7, 9731.7, 24450.34, -13594, 546.3613,-5.11122,-0.13944, 8.075055, 8.60821, 1.422208,

3.2. Hasil Deteksi Random Forest

Setelah melalui proses pemilihan fitur, data dibagi menjadi dua bagian yaitu data latih dan data uji. Langkah pertama adalah mempelajari cara menemukan bobot dan bias hierarki jaringan *random forest*. Hasil pembelajaran *random forest* adalah bobot dan bias yang digunakan untuk membangun jaringan IoT yang kompleks. Berikut proses pendeteksian yang menggunakan algoritma *random forest* untuk mendeteksi serangan pada jaringan IoT. Selanjutnya, sistem deteksi intrusi (IDS) diuji pada jaringan IoT menggunakan data input yang fiturnya diekstraksi menggunakan PCA. Hasil percobaan menunjukkan bahwa Random Forest mampu mendeteksi serangan pada catatan TCP dengan tingkat keberhasilan yang relatif tinggi dan jumlah kesalahan yang terdeteksi sedikit.

Tabel 3. Hasil pengujian deteksi Random forest

Pengujian	Jumlah Fitur	Akurasi	Presisi	Recall
1	5	0.9919	0.6358	0.4833
	8	0.9928	0.5115	0.4368
	10	0.9931	0.6387	0.4779
	All	0.9934	0.5184	0.4370
2	5	0.9919	0.6358	0.4833
	8	0.9928	0.5115	0.4368
	10	0.9931	0.6387	0.4779
	All	0.9934	0.5184	0.4370
3	5	0.9919	0.6358	0.4833
	8	0.9928	0.5115	0.4368
	10	0.9931	0.6387	0.4779
	All	0.9934	0.5184	0.4370
4	5	0.9919	0.6358	0.4833
	8	0.9928	0.5115	0.4368
	10	0.9931	0.6387	0.4779
	All	0.9934	0.5184	0.4370
5	5	0.9919	0.6358	0.4833
	8	0.9928	0.5115	0.4368
	10	0.9931	0.6387	0.4779
	All	0.9934	0.5184	0.4370

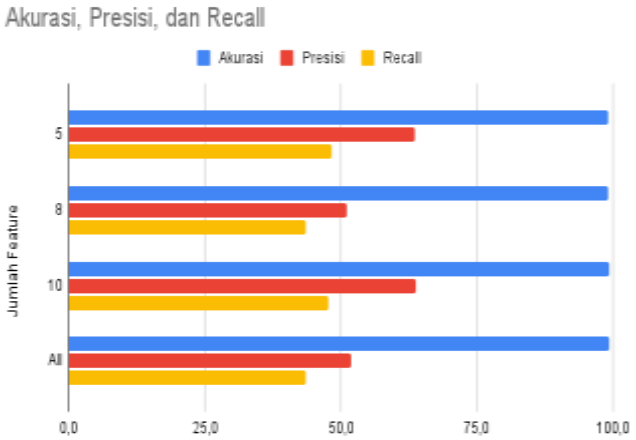

Fig.3. Hasil Pengujian Akurasi, Presisi dan Recall

Gambar 3 adalah hasil pengujian menggunakan algoritma Random Forest, pada IDS jaringan CIC IoT 2023 menunjukkan hasil yang sangat memuaskan dengan tiga parameter evaluasi utama: akurasi, presisi, dan recall. Hasil akurasi tertinggi diperoleh pada pengujian ke-5 dengan jumlah 5 fitur. Rata-rata akurasi pengujian deteksi menggunakan algoritma Random Forest pada jaringan kompleks IoT mencapai 99.2%

4. Conclusion

Perkembangan teknologi saat ini mempengaruhi kerentanan keamanan pada jaringan IoT. Penelitian ini mengusulkan sistem deteksi menggunakan metode hybrid *PCA-Random Forest*. Pada penelitian ini dataset yang digunakan adalah CIC IoT 2023 yang mencakup 47 fitur. Fitur-fitur tersebut kemudian diseleksi menggunakan algoritma Principal Component Analysis (PCA), sehingga diperoleh seleksi sebanyak 5, 8, dan 10 fitur, dan tanpa PCA diperoleh 47 fitur. Dari dataset tersebut dilakukan data pelatihan dan pengujian menggunakan model machine learning khususnya *Random Forest Classifier*. Dari penelitian ini dapat disimpulkan bahwa nilai akurasinya sebesar 99,2%. Hasil dengan akurasi tertinggi diperoleh pada pengujian ke-5 dengan 5 fitur. Hasil tersebut menunjukkan bahwa penggunaan metode *Random Forest Classification* dan PCA dapat meningkatkan akurasi deteksi jaringan data CIC IoT 2023.

Acknowledgment

Penelitian ini didukung oleh Universitas Dinamika Bangsa, Jambi, Indonesia

References

[1] I. Sumaiya Thaseen, J. Saira Banu, K. Lavanya, M. Rukunuddin Ghalib, and K. Abhishek, "An integrated intrusion detection system using correlation-based attribute selection and artificial neural network," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 2, pp. 1–15, 2021, doi: 10.1002/ett.4014.

[2] V. No, J. Hal, F. Adi, R. Anggi, D. Puji, and E. Kartikadarma, "Optimasi Algoritma Random Forest menggunakan Principal Component Analysis untuk Deteksi Malware," vol. 5, no. 3, pp. 217–223, 2023.

[3] A. Nugraha and N. Rijati, "Penerapan Metode Principal Component Analysis (PCA) Untuk Deteksi Anomali Pada Jaringan Peer-To-Peer (P2P) Botnet," *Techno.COM*, vol. 14, no. 3, pp. 212–217, 2015.

[4] R. Pangestu and A. Solichin, "Klasifikasi Serangan Jaringan Menggunakan Metode Decision Tree Berbasis Website," *Semin. Nas. Mhs. Fak. Teknol. Inf.*, no. September, pp. 614–620, 2022,

[5] T. Oshiro, P. Perez, and J. Baranauskas, "How Many Trees in a Random Forest?," in *Lecture notes in computer science*, 2012, vol. 7376, doi: 10.1007/978-3-642-31537-4_13.

[6] R. L. Atimi and Enda Esyudha Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *J. Sains dan Inform.*, vol. 8, no. 1, pp. 88–96, 2022, doi: 10.34128/jsi.v8i1.419.

[7] A. Fathan Hidayatullah and A. Sn, "Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter," *Semin. Nas. Inform.*, vol. 2017, no. semnasIF, pp. 115–122, 2017.

[8] R. Sudiyarno, A. Setyanto, and E. T. Luthfi, "Peningkatan Performa Pendeteksian Anomali Menggunakan Ensemble Learning dan Feature Selection Anomaly Detection Performance Improvement Using Ensemble Learning and Feature Selection," *Citec J.*, vol. 7, no. 1, pp. 1–9, 2020.

[9] E. Carlos *et al.*, "CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment," 2023.

[10] R. Rahul, "Random Forest Classification and it's Mathematical Implementation. Medium.," 2020.

[11] M. F. R. Alif P. B.A, Yudha P., "Deteksi Malware Menggunakan Metode Random Forest Berdasarkan Analisa Forensik," 2021.

Authors' Profiles

Adrian Pirtama lahir di Jambi, Indonesia. Saat ini dia tengah menempuh Program Sarjanah Informatika di Universitas Dinamika Bangsa, Indonesia. Dia memiliki fokus penelitian pada machine learning dan software engineering.

Yuda Prasetia lahir di Jambi, Indonesia. Saat ini dia tengah menempuh Program Sarjanah Informatika di Universitas Dinamika Bangsa, Indonesia. Dia memiliki fokus penelitian pada machine learning dan software engineering.

Redho Irnindo Saputra lahir di Jambi, Indonesia. Saat ini dia tengah menempuh Program Sarjanah Informatika di Universitas Dinamika Bangsa, Indonesia. Dia memiliki fokus penelitian pada machine learning dan software engineering

Eko Arip Winanto meraih gelar Sarjana di bidang Sistem Komputer dari Universitas Sriwijaya, Indonesia, dan gelar M.Phil di bidang Ilmu Komputer dari Universiti Teknologi Malaysia, Malaysia. Saat ini beliau adalah dosen di Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Indonesia. Minat penelitiannya meliputi IoT, pembelajaran mesin, blockchain, dan keamanan jaringan.