Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) DOI: 10.25126/jtiik.20241127678
Vol. 11, No. 2, April 2024, hlm. 285-290 p-ISSN: 2355-7699
Akreditasi KEMENRISTEKDIKTI, No. 36/E/KPT/2019 e-ISSN: 2528-6579

PENINGKATAN PERFORMA DETEKSI SERANGAN MENGGUNAKAN METODE PCA DAN RANDOM FOREST

Eko Arip Winanto[1], Yudi Novianto [*2], Sharipuddin[3], Ibnu Sani Wijaya[4], Pareza Alam Jusia[5]

[1,2,3,4,5] Universitas Dinamika Bangsa, Surabaya
Email: [1]ekoaripwinanto@unama.ac.id, [2]viant.yudi@gmail.com, [3]sharipuddin@unama.ac.id, [4]ibnu_sw17@unama.ac.id., [5]parezaalam@gmail.com
[*]Penulis Korespondensi

(Naskah masuk: 23 Agustus 2023, diterima untuk diterbitkan: 04 April 2024)

Abstrak

Keamanan jaringan menjadi hal yang sangat penting dalam menghadapi ancaman serangan yang semakin kompleks dan canggih. Deteksi serangan dalam jaringan dapat membantu mengidentifikasi aktivitas mencurigakan yang mengindikasikan upaya penetrasi atau serangan oleh pihak yang tidak berwenang. Dalam upaya untuk meningkatkan performa deteksi serangan pada jaringan IoT perlu adanaya penerapan sebuah metode untuk mendeteksi sebuah ancaman . Metode *Random Forest* adalah algoritma pembelajaran mesin yang memanfaatkan ansambel pohon keputusan. Ansambel tersebut terdiri dari beberapa pohon keputusan independen yang digunakan untuk mengklasifikasikan data. Salah satu karakteristik dari metode *Random Forest* adalah kemampuannya dalam mengatasi masalah overfitting dan kualitas prediksi yang baik. *Principal Component Analysis* (PCA) adalah teknik statistik yang digunakan untuk mengurangi dimensi data dengan memproyeksikannya ke ruang fitur yang lebih rendah. Hal ini membantu menghilangkan korelasi antar fitur dan mengidentifikasi fitur-fitur penting yang dapat meningkatkan pemisahan antara serangan dan lalu lintas normal. Dalam penelitian ini akan diujikan dengan dataset CIC IOT 2023 yang terdiri dari beberapa tipe serangan yaitu *DDoS, DoS, Recon, Web-based, Brute Force, Spoofing, dan Mirai*. Pengujian model terdiri dari 4 fitur yaitu 5,8,10 dan 47. Hasil deteksi menunjukkan hasil yang memuaskan dengan meningkatkan kinerja dalam mendeteksi serangan hingga mencapai 99,2%

Kata kunci: *Intrusion Detection, Machine Learning, Network Security, Principal Component Analysis, Random Forest, Threat Detection*

ENHANCEMENT ATTACK DETECTION USING PCA AND RANDOM FOREST METHOD

Abstract

Network security has become increasingly critical in the face of complex and sophisticated threat attacks. Detecting intrusions within a network can aid in identifying suspicious activities indicative of unauthorized penetration attempts or attacks. To enhance intrusion detection performance, the implementation of a method for threat detection is necessary. The Random Forest method, an ensemble machine learning algorithm that leverages multiple independent decision trees, is employed in this study. This method effectively addresses overfitting issues and demonstrates good predictive quality. Principal Component Analysis (PCA), a statistical technique for dimensionality reduction, is utilized to project data into a lower-dimensional feature space. By eliminating correlations between features and identifying important ones, PCA enhances the separation between attacks and normal traffic. This research utilizes the CIC IOT 2023 dataset, encompassing various types of attacks such as DDoS, DoS, Recon, Web-based, Brute Force, Spoofing, dan Mirai. The model testing phase incorporates 4 features: 5, 8, 10, and 47. The detection results indicate a remarkable performance improvement in identifying attacks, achieving an accuracy rate of 99.2%.

Keywords: *Intrusion Detection, Machine Learning, Network Security, Principal Component Analysis, Random Forest, Threat Detection*

1. PENDAHULUAN

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. 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). Penelitian ini akan fokus pada penerapan IDS untuk mengidentifikasi komunikasi data yang mencurigakan atau tidak normal (Sumaiya Thaseen et al., 2021). Tantangan pada sistem IDS adalah bagaimana mengusulkan sebuah metode dengan tingkat akurasi deteksi yang tinggi. Beberapa penelitian (Sudiyarno, Setyanto and Luthfi, 2020) (Nugraha and Rijati, 2015) telah menggusulkan metode untuk peningkatan sistem deteksi.

Pada penelitian (Sudiyarno, Setyanto and Luthfi, 2020) menunjukan analisis penelitian yaitu metode *Ensemble learning* dan *feature selection.* Hasil penelitian tersebut menunjukkan bahwa pendekatan tersebut efektif dalam mendeteksi intrusi anomali. Selain itu, penelitian ini mencapai tingkat akurasi yang tinggi yaitu sebesar 94,5%. Kemudian, mengusulkan penggunaan feature selection untuk mengurangi waktu eksekusi secara signifikan.

Penelitian lainya menggusulkan optimasi algoritma *random forest* menggunakan *principal component* analysis untuk deteksi malware (Nugraha and Rijati, 2015). Penelitian ini membahas tentang peningkatan performa algoritma Random Forest menggunakan PCA telah selesai dilakukan. *Random Forest* dipilih karena memiliki performa Akurasi dari *Recall* terbaik dibandingkan empat algoritma lain, seperti: *Adaboost*, *Neural Network*, *Support Vector Machine* dan *kNearest Neighbor*. Oleh karena itu pada penelitian ini akan mengusulkan metode *random forest* untuk metode deteksi.

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 (Oshiro, Perez and Baranauskas, 2012). Sehingga memiliki potensi untuk digunakan sebagai metode deteksi serangan. Pada penelitian (Adi et al., 2023) mengusulkan penerapan metode *feature extraction* untuk proses pemilihan fitur.

Tujuan dari *feature extraction* adalah untuk meningkatkan akurasi dari sistem deteksi serangan. Salah satu metode feature extraction adalah PCA digunakan untuk mencari rekomendasi faktor-faktor yang mempengaruhi serangan pada jaringan. PCA adalah teknik untuk mengekstraksi struktur dari suatu dataset. Pemilihan beberapa faktor spesifik menggunakan algoritma PCA dimaksudkan untuk menciptakan teknik deteksi yang memiliki akurasi yang baik.

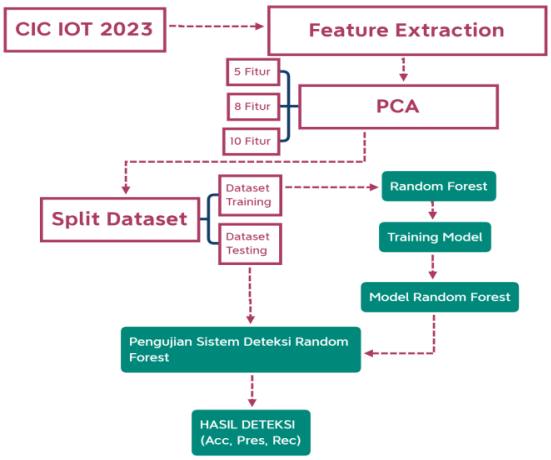
Oleh karena itu pada penelitian ini menggusulkan metode hibrid *Random Forest Classifier* untuk metode deteksi dan metode PCA sebagai ekstraksor factor yang digunakan untuk menghasilkan deteksi yang ideal (Atimi and Enda Esyudha Pratama, 2022) (Widowati and Sadikin, 2021). Untuk susunan selanjutnya pada penelitian ini terdiri dari bagian 2 berisi metode penelitian. Bagian ketiga adalah hasil dari pengujian dan 3 berisi pembahasan atau diskusi. Terakhir berisi kesimpulan akir dari penelitian ini.

2. METODE PENELITIAN

2.1 Experiment Setup

Penelitian ini bertujuan untuk mendeteksi serangan pada jaringan. Pada penelitian ini menggunakan metode *Random Forest* untuk mendeteksi serangan tersebut. Ada beberapa tahapan yang perlu dilakukan pada penelitian ini. Gambar 1 merupakan konfigurasi eksperimen yang dirancang untuk penelitian ini. Penelitian ini dibagi menjadi tiga tahapan yang berbeda.

- Pada tahap pertama, extraksi fitur dilakukan dari dataset dengan menggunakan metode PCA serta dataset dibagi menjadi dataset training dan dataset testing.
- Pada tahap kedua, proses training model dilakukan dengan menggunakan dataset training sehinga diperoleh hasil model *Random Forest* untuk deteksi serangan pada jaringan.
- Pada tahap terakhir, dilakukan tahap pengujian dengan menggunakan *dataset testing* pada model yang telah di hasilkan pada tahap sebelumnya untuk menghitung tingkat keberhasilan seperti akurasi, presisi dan *recall*



Gambar 1. Experiment setup

2.2 Dataset

Penelitian ini akan menggunakan dataset CIC IoT 2023. Pada dataset ini Serangan diklasifikasikan ke dalam tujuh kategori, yaitu DDoS, DoS, Recon, Web-based, Brute Force, Spoofing, dan Mirai. Terakhir, semua serangan dilakukan oleh perangkat

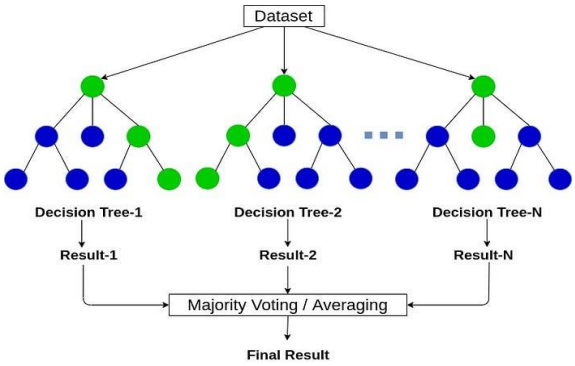
IoT jahat yang menargetkan perangkat IoT lainnya (Neto et al., 2023). Pengunaan dataset ini dikarenakan memiliki 7 tipe kelas serangan pada jaringan IoT dan 33 jenis serangan.

2.3 PCA

Ekstraksi fitur adalah proses yang berguna dalam mengidentifikasi fitur-fitur penting dari data, sehingga membuatnya lebih cepat, lebih mudah, dan lebih dapat dipahami. Proses ekstraksi fitur memiliki dampak yang signifikan terhadap kualitas klasifikasi. Akurasi dari Sistem Deteksi Intrusi (IDS) akan berbeda-beda tergantung pada beragamnya fitur-fitur yang digunakan sebagai input. Selain itu, hasil dari proses klasifikasi juga dipengaruhi oleh lalu lintas yang padat dalam jaringan yang kompleks, seperti pada Internet of Things, serta penggunaan fitur-fitur multidimensi (Lee, Pak and Lee, 2020). Dalam penelitian ini, digunakan Analisis Komponen Utama (PCA) sebagai metode untuk mengurangi dimensi dari dataset (Sharipuddin et al., 2023) (Sharipuddin et al., 2020). Penelitian ini akan menggunakan dataset dengan dimensi 5, 8, 10, dan 47 fitur, yang akan digunakan dalam proses pelatihan dan deteksi

2.4 Random Forest

Algoritma Random Forest memiliki kemampuan untuk melakukan klasifikasi pada data yang memiliki atribut yang tidak lengkap dan sesuai digunakan untuk pengklasifikasian data sampel yang besar. Dalam proses klasifikasi Random Forest, data sampel akan dibagi secara acak ke dalam decision tree (Wanli Sitorus, Sukarno and Mandala, 2021) (Chen and Chen, 2020). Setelah pembentukan tree, setiap tree memiliki node root (akar), node internal (cabang-cabang), dan node leaf (hasil kelas) (Jin et al., 2020) (Jmila and Khedher, 2022). Karateristik metode RF dapat di lihat pada Gambar 2. Untuk penggunaan metode RF pada sistem deteksi yang diusulkan pada penelitian ini dapat dilihat pada Tabel 1.



Gambar 2. Arsitektur RF

Tabel 1. Metode RF-IDS

Pseudocode RF-IDS
dataset → CICIoT2023 *X = [sample[:-1] for sample in dataset]* *y = [sample[-1] for sample in dataset]* *X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)* *num_trees → 10* *max_depth → 5* *num_features →4* *random_forest→RandomForestClassifier(n_estimators=num_trees, max_depth=max_depth, max_features=num_features, random_state=42)* *random_forest.fit(X_train, y_train)* *y_pred → random_forest.predict(X_test)* *accuracy → accuracy_score(y_test, y_pred)*

2.5 Environment Setup

Semua experiment yang dilakukan dalam penelitian ini, dilakukan dengan menggunakan laptop yang memiliki spesifikasi sebagai berikut: Sistem Operasi: *Windows 10 Pro 64-bit (10.0, Build 19045) Processor: AMD Ryzen 5 PRO 2500U w/ Radeon Vega Mobile Gfx (8 CPUs), ~2.0 GHz RAM: 16 GB*. Adapun tools yang diperlukan untuk melakukan analisis adalah *Python*, *Scikit-Learn*, *TensorFlow* dan *Keras*

3. HASIL DAN PEMBAHASAN

Bagian ini merupakan hasil eksperimen yang dilakukan, termasuk hasil dari proses reduksi fitur dan pengujian kinerja algoritma *Random Forest*. Pembahasan meliputi evaluasi reduksi fitur yang dilakukan dan analisis kinerja algoritma *Random Forest*

3.1 Hasil PCA

Pada fitur ekstraksi ini algoritma Random Forest akan ditingkatkan dengan menerapkan metode Principal Componnent Analysis (PCA) untuk reduksi jumlah fiturnya. Tujuannya adalah untuk mengurangi upaya komputasi dan meningkatkan kinerja sistem identifikasi dalam jaringan IoT tanpa menghilangkan karakteristik pada datanya. Tabel 2 adalah hasil dari ekstraksi fitur dari fitur semula menjadi tiga kategori yaitu 5, 8 dan 10 fitur. Hasil dari PCA menjadi sebuah nilai yang berupa nilai angka. Fitur ini mewakili nilai dari fitur sebelumnya yang tanpa kehilangan ciri dari fitur sebelumnya.

Tabel 2. Hasil PCA

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,

Hasil reduksi fitur ini digunakan untuk data pelatihan IDS menggunakan algoritma *Random Forest.*

3.2 Hasil Sistem Deteksi

Setelah melalui proses seleksi fitur, data tersebut dibagi menjadi dua bagian, yaitu data pelatihan (training) dan data pengujian (testing). Langkah pertama adalah mempelajari cara menemukan model dari Random Forest. Hasil pembelajaran RF adalah model yang digunakan untuk membangun sistem IDS.

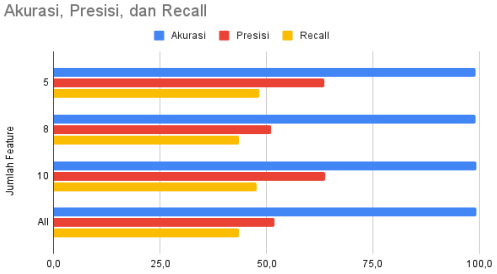
Selanjutnya, dilakukan pengujian sistem deteksi intrusi (IDS) pada jaringan IDS-IoT dengan menggunakan data input yang telah diekstraksi fiturnya menggunakan PCA. Hasil pengujian menunjukkan bahwa Random Forest mampu mendeteksi serangan pada dengan tingkat keberhasilan yang relatif tinggi dan jumlah kesalahan deteksi yang sedikit.

Tabel 3 adalah hasil dari pengujian deteksi menggunakan RF. Dari hasil pengujian ini diperolah bahwa hasil pengujian menunjukan keberhasilan dalam mendeteksi serangan yang terjadi. Hasil akurasi terbaik diperolah pada fitur yang berjumlah 10 fitur.

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	0.9234	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	0.8934	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	0.904	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	0.9234	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	0.9134	0.5184	0.4370

Selanjutnya adalah pembahasan dari hasil pengujian sistem deteksi menggunkan RF. Pada 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 fitur dengan jumlah 10 fitur.



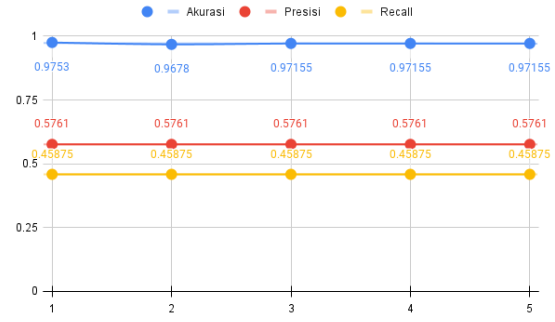
Gambar 3. Hasil perbandingan akuasi, presisi dan recall

4. DISKUSI

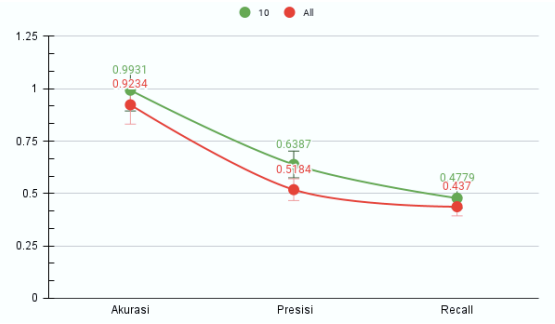
Hasil dari pengujian deteksi serangan pada dataset yang diujikan berhasil mendeteksi dengan memuaskan. Pada tabel 4 dan Gambar 4 adalah hasil rata-rata pengujian dari setiap pengujian. Hasil rata-rata menunjukkan hasil akurasi tertinggi mencapai 97.5% dan rata-rata terendah mencapai 96%.

Tabel 4. Hasil pengujian deteksi Random forest

Pengujian	Akurasi	Presisi	Recall
1	0.9753	0.5761	0.45875
2	0.9678	0.5761	0.45875
3	0.97155	0.5761	0.45875
4	0.97155	0.5761	0.45875
5	0.97155	0.5761	0.45875



Gambar 4. Hasil rata-rata perbandingan akuasi, presisi dan recall



Gambar 5. Hasil perbandingan akuasi, presisi dan recall 10 fitur dan semua fitur

Selanjutnya pada gambar 5 adalah perbandingan peforma dari pengujian deteksi menggunakan algoritma Random Forest pada jaringan kompleks IoT tertinggi pada fitur 10 mencapai 99.2% dan terendah pada fitur semua. Hasil pengujian yang kurang memuaskan terdapat pada pengujian pada semua fitur. Hasil ini menunjukan bahwa penggunaan PCA pada penelitian ini berhasil meningkatkan akurasi RF tanpa menggunakan metode PCA.

5. KESIMPULAN

Perkembangan teknologi saat ini mempengaruhi kerentanan keamanan pada jaringan IoT. Penelitian ini mengusulkan sistem deteksi menggunakan hybrid Pca- Random Forest. Pada penelitian ini *dataset* yang digunakan adalah CIC IoT 2023 yang terdiri dari 47 fitur dan serangan seperti DDoS, DoS, Recon, Web-based, Brute Force,

Spoofing, dan Mirai. Fitur-fitur tersebut kemudian diseleksi menggunakan algoritma *Principal Component Analysis* (PCA), menghasilkan 5, 8, dan 10 fitur yang dipilih serta tanpa PCA yaitu 47 fitur. Dari dataset tersebut dilakukan training dan testing data dengan menggunakan model *machine learning* yaitu *Random Forest Classifier*. Dari penelitian ini dapat disimpulkan menujukkan bahwan hasil nilai akurasi sebesar 99.2%. Hasil akurasi tertinggi diperoleh pada jumlah 10 fitur. Hasil ini menunjukkan bawah pengguanan metode *Random Forest Classifier* dan PCA dapat meningkatkan akurasi deteksi dari pada jaringan dataset CIC IoT 2023

DAFTAR PUSTAKA

ADI, F., ANGGI, R., PUJI, D. AND KARTIKADARMA, E., 2023. Optimasi Algoritma Random Forest menggunakan Principal Component Analysis untuk Deteksi Malware. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(3), pp.217–223.

ATIMI, R.L. AND ENDA ESYUDHA PRATAMA, 2022. Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia. *Jurnal Sains dan Informatika*, 8(1), pp.88–96. https://doi.org/10.34128/jsi.v8i1.41.

CHEN, M.M. AND CHEN, M.C., 2020. Modeling road accident severity with comparisons of logistic regression, decision tree and random forest. *Information (Switzerland)*, 11(5). https://doi.org/10.3390/INFO11050270.

JIN, D., LU, Y., QIN, J., CHENG, Z. AND MAO, Z., 2020. SwiftIDS: Real-time intrusion detection system based on LightGBM and parallel intrusion detection mechanism. *Computers and Security*, [online] 97, p.101984. https://doi.org/10.1016/j.cose.2020.101984.

JMILA, H. AND KHEDHER, M.I., 2022. Adversarial machine learning for network intrusion detection: A comparative study. *Computer Networks*, [online] 214(May), p.109073. https://doi.org/10.1016/j.comnet.2022.109073.

KHAN, M.A., 2021. HCRNNIDS : Hybrid Convolutional Recurrent Neural. *Multidisciplinary Digital Publishing Institute*, 8(834).

LEE, J., PAK, J.G. AND LEE, M., 2020. Network Intrusion Detection System using Feature Extraction based on Deep Sparse Autoencoder. *International Conference on ICT Convergence*, 2020-October, pp.1282–1287. https://doi.org/10.1109/ICTC49870.2020.9289253.

NETO, E.C.P., DADKHAH, S., FERREIRA, R., ZOHOURIAN, A., LU, R. AND GHORBANI, A.A., 2023. CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment. *Sensors*, 23(13), p.5941. https://doi.org/10.3390/s23135941.

NUGRAHA, A. AND RIJATI, N., 2015. Penerapan Metode Principal Component Analysis (PCA) Untuk Deteksi Anomali Pada Jaringan Peer-To-Peer (P2P) Botnet. *Techno.COM*, 14(3), pp.212–217.

OSHIRO, T.M., PEREZ, P.S. AND BARANAUSKAS, J.A., 2012. How many trees in a random forest? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7376 LNAI, pp.154–168. https://doi.org/10.1007/978-3-642-31537-4_13.

SHARIPUDDIN, PURNAMA, B., KURNIABUDI, WINANTO, E.A., STIAWAN, D., HANAPI, D., IDRIS, M.Y. BIN AND BUDIARTO, R., 2020. Features extraction on iot intrusion detection system using principal components analysis (Pca). *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2020-Octob, pp.114–118. https://doi.org/10.23919/EECSI50503.2020.9251292.

SHARIPUDDIN, WINANTO, E.A., MOHTAR, Z.Z., KURNIABUDI, WIJAYA, I.S. AND SANDRA, D., 2023. Improvement detection system on complex network using hybrid deep belief network and selection features. *Indonesian Journal of Electrical Engineering and Computer Science*, 31(1), pp.470–479. https://doi.org/10.11591/ijeecs.v31.i1.pp470-479.

SUDIYARNO, R., SETYANTO, A. AND LUTHFI, E.T., 2020. Peningkatan Performa Pendeteksian Anomali Menggunakan Ensemble Learning dan Feature Selection Anomaly Detection Performance Improvement Using Ensemble Learning and Feature Selection. *Citec Journal*, 7(1), pp.1–9.

SUMAIYA THASEEN, I., SAIRA BANU, J., LAVANYA, K., RUKUNUDDIN GHALIB, M. AND ABHISHEK, K., 2021. An integrated intrusion detection system using correlation-based attribute selection and artificial neural network. *Transactions on Emerging Telecommunications Technologies*, 32(2), pp.1–15. https://doi.org/10.1002/ett.4014.

WANLI SITORUS, Y., SUKARNO, P. AND MANDALA, S., 2021. Analisis Deteksi Malware Android menggunakan metode Support Vector Machine & Random Forest. *e-Proceeding of Engineering*, 8(6), pp.12500–12518.

Halaman ini sengaja dikosongkan.