Jurnal Minfo Polgan
Volume 13, Nomor 2, September 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i2.14074 p-ISSN : 2089-9424

Terbit : 02 September 2024

Implementasi Machine Learning untuk Deteksi Intrusi pada Jaringan Komputer

[1]Dyan Prawita Sari, [2] Zuhri Halim, [3] Irlon, [4] Bayu Waseso, [5] Saromah
[1] Universitas Gunadarma, Indonesia, [2] Universitas Muhammadiyah Prof Dr Hamka, [3] Institut Teknologi Budi Utomo, [4] Universitas Mercu Buana, [5] Universitas Indraprastha PGRI

[1] wprawita@staff.gunadarma.ac.id, [2] zuhri@uhamka.ac.id, [3] irlon@itbu.ac.id,
[4]bayu.waseso@mercubuana.ac.id, [5] saromah73@gmail.com

ABSTRAK

Dalam era digital yang semakin berkembang, keamanan jaringan komputer menjadi isu yang sangat penting, terutama dengan meningkatnya ancaman dari serangan siber. Salah satu metode yang efektif dalam mendeteksi ancaman tersebut adalah melalui implementasi machine learning. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model machine learning yang mampu mendeteksi intrusi pada jaringan komputer secara real-time. Model yang diusulkan menggunakan teknik supervised learning, di mana dataset yang berisi lalu lintas jaringan normal dan lalu lintas yang mengandung serangan digunakan untuk melatih algoritma. Algoritma yang dipertimbangkan meliputi Decision Tree, Random Forest, dan Support Vector Machine (SVM). Penelitian ini juga melakukan analisis komparatif untuk menilai kinerja masing-masing algoritma dalam hal akurasi, presisi, recall, dan waktu pemrosesan. Hasil eksperimen menunjukkan bahwa model machine learning yang diterapkan mampu mendeteksi berbagai jenis serangan dengan tingkat akurasi yang tinggi, mencapai lebih dari 95% pada dataset uji. Selain itu, Random Forest terbukti menjadi algoritma yang paling efektif dalam mendeteksi intrusi dengan keseimbangan terbaik antara akurasi dan waktu pemrosesan. Implementasi sistem ini diharapkan dapat meningkatkan kemampuan deteksi intrusi pada jaringan komputer, sehingga membantu dalam menjaga keamanan data dan mengurangi potensi kerugian akibat serangan siber.

Kata Kunci: Machine Learning, Deteksi Intrusi, Jaringan Komputer

PENDAHULUAN

Di era digital saat ini, jaringan komputer memainkan peran vital dalam hampir setiap aspek kehidupan, mulai dari bisnis, pendidikan, hingga pemerintahan. Namun, seiring dengan semakin kompleks dan terintegrasinya infrastruktur jaringan, ancaman terhadap keamanan siber juga meningkat secara signifikan. Serangan siber, seperti Distributed Denial of Service (DDoS), malware, dan intrusi jaringan, dapat menyebabkan kerugian yang sangat besar, baik dari segi finansial maupun reputasi. Oleh karena itu, penting untuk memiliki sistem deteksi intrusi (Intrusion Detection System/IDS) yang efektif dan responsif. (Maxwell et al., 2018)

Sistem deteksi intrusi tradisional, yang umumnya berbasis pada aturan (rule-based), memiliki keterbatasan dalam mendeteksi serangan yang belum pernah terjadi sebelumnya atau serangan yang menggunakan teknik baru. Di sinilah machine learning menawarkan solusi yang lebih adaptif dan canggih. Dengan kemampuan untuk belajar dari data sebelumnya dan mengenali pola yang mencurigakan, machine learning dapat secara signifikan meningkatkan efektivitas deteksi intrusi. (Bororing, 2024)

Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi model machine learning dalam mendeteksi intrusi pada jaringan komputer. Fokus utama dari penelitian ini adalah mengidentifikasi algoritma machine learning yang paling efektif dalam mendeteksi berbagai jenis serangan siber. Algoritma yang digunakan dalam penelitian ini meliputi Decision Tree, Random Forest, dan Support Vector Machine (SVM). Selain itu, penelitian ini juga akan mengkaji performa model berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan waktu pemrosesan.

Jurnal Minfo Polgan
Volume 13, Nomor 2, September 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i2.14074 p-ISSN : 2089-9424

(Heizmann et al., 2022)

Dengan meningkatnya ancaman siber dan kebutuhan akan sistem keamanan yang lebih canggih, penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan teknologi deteksi intrusi yang lebih efektif dan efisien, serta memberikan landasan bagi pengembangan IDS berbasis machine learning di masa mendatang. (Xie et al., 2020)

TINJAUAN PUSTAKA

Keamanan Jaringan Komputer: Konsep dan Tantangan Keamanan jaringan komputer adalah bidang yang berfokus pada perlindungan integritas, kerahasiaan, dan ketersediaan data yang dikirimkan melalui jaringan. Ancaman terhadap jaringan komputer mencakup berbagai jenis serangan, seperti serangan Distributed Denial of Service (DDoS), malware, sniffing, dan berbagai bentuk intrusi lainnya. Menurut (Alotaibi, 2019) keamanan jaringan melibatkan penggunaan berbagai teknik dan alat untuk mengidentifikasi, mencegah, dan memitigasi ancaman ini. Namun, kompleksitas jaringan modern, yang melibatkan berbagai perangkat dan protokol, membuat deteksi ancaman menjadi semakin menantang. (Char et al., 2018)

Prinsip dan Klasifikasi Intrusion Detection System (IDS): Intrusion Detection System (IDS) terdiri dari dua kategori utama: Network-based Detection System (NIDS) dan Host-based Detection System (HIDS). NIDS melacak lalu lintas jaringan untuk mendeteksi ancaman, sementara HIDS melacak aktivitas pada perangkat individu. Mekanisme deteksi intrusi (IDS) konvensional biasanya bergantung pada aturan yang telah ditentukan sebelumnya untuk mendeteksi serangan. Namun, telah terbukti bahwa teknik ini tidak efektif dalam menangani serangan baru atau variasi dari serangan sebelumnya.

Peran Machine Learning dalam Deteksi Intrusi Machine learning telah menjadi salah satu pendekatan yang paling menjanjikan dalam deteksi intrusi, karena kemampuannya untuk mengenali pola dari data historis dan mendeteksi anomali secara otomatis. Menurut (Singh et al., 2021), machine learning dapat diklasifikasikan menjadi tiga jenis utama: supervised learning, unsupervised learning, dan reinforcement learning. Supervised learning memerlukan data yang telah dilabeli untuk melatih model, sedangkan unsupervised learning berfokus pada pengenalan pola dalam data yang tidak berlabel. Dalam konteks deteksi intrusi, supervised learning lebih umum digunakan, terutama dengan algoritma seperti Decision Tree, Random Forest, dan Support Vector Machine (SVM), yang telah terbukti efektif dalam berbagai studi. (Singh et al., 2021)

Algoritma Machine Learning untuk Deteksi Intrusi Beberapa algoritma machine learning telah diterapkan dalam deteksi intrusi, masing-masing dengan kelebihan dan kekurangan tersendiri:

a. Decision Tree: Algoritma ini bekerja dengan membagi dataset menjadi subset berdasarkan fitur yang paling informatif, yang menghasilkan model berbentuk pohon keputusan. Decision Tree populer karena interpretabilitasnya yang tinggi, tetapi cenderung overfitting pada data yang kompleks. (Bhanu et al., 2020)
b. Random Forest: Merupakan ensemble dari banyak pohon keputusan, Random Forest mengurangi risiko overfitting dengan menggabungkan prediksi dari beberapa pohon, sehingga meningkatkan akurasi, menunjukkan bahwa Random Forest dapat memberikan hasil yang sangat akurat dalam berbagai aplikasi, termasuk deteksi intrusi.
c. Support Vector Machine (SVM): SVM bekerja dengan mencari hyperplane yang memisahkan data dari dua kelas dengan margin maksimum. SVM telah digunakan secara luas dalam deteksi intrusi karena kemampuannya menangani data non-linear dengan menggunakan kernel trick.

Dataset dan Evaluasi Model Machine Learning Pemilihan dataset yang tepat merupakan faktor kunci dalam pengembangan model machine learning untuk deteksi intrusi. Dataset yang umum digunakan dalam penelitian ini adalah KDD Cup 99, NSL-KDD, dan CICIDS 2017. Meacham et al. (2009) menunjukkan bahwa kualitas dataset sangat mempengaruhi performa model, karena dataset yang tidak representatif dapat menyebabkan model yang dibangun memiliki bias atau tidak mampu mendeteksi intrusi dengan efektif. Metrik evaluasi yang sering digunakan dalam penelitian ini meliputi akurasi, presisi, recall, dan F1-score, yang memberikan gambaran tentang kinerja model dalam mendeteksi ancaman dengan tepat. (Mall et al., 2022)

Tantangan dan Arah Penelitian Masa Depan Meskipun telah ada banyak kemajuan dalam

Jurnal Minfo Polgan
Volume 13, Nomor 2, September 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i2.14074 p-ISSN : 2089-9424

penggunaan machine learning untuk deteksi intrusi, masih ada beberapa tantangan yang perlu diatasi. Salah satunya adalah kebutuhan untuk mengembangkan model yang dapat beradaptasi dengan cepat terhadap ancaman baru dan tidak terdeteksi (zero-day attacks). Selain itu, pengurangan false positives tetap menjadi fokus penting, karena false positives yang tinggi dapat mengurangi kepercayaan pengguna terhadap sistem IDS. Penelitian masa depan juga diharapkan dapat mengeksplorasi integrasi machine learning dengan teknologi baru, seperti blockchain dan edge computing, untuk meningkatkan keamanan jaringan secara keseluruhan.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan tujuan untuk mengembangkan dan mengevaluasi model machine learning dalam mendeteksi intrusi pada jaringan komputer. Metode penelitian ini dibagi menjadi beberapa tahap, yang mencakup pengumpulan data, pemrosesan data, pemilihan model, pelatihan dan pengujian model, serta evaluasi kinerja model. Penjelasan rinci dari setiap tahap adalah sebagai berikut:

1. Pengumpulan Data

Penelitian ini menggunakan dataset NSL-KDD, yang merupakan versi yang lebih bersih dan lebih representatif dari dataset KDD Cup 99. Dataset ini dipilih karena mencakup berbagai jenis serangan jaringan serta lalu lintas normal, yang memungkinkan model untuk dilatih pada skenario yang realistis. Dataset NSL-KDD terdiri dari 41 fitur yang mencakup informasi seperti durasi koneksi, jumlah paket yang dikirim, dan jenis protokol yang digunakan. Dataset ini terdiri dari dua bagian utama: data pelatihan dan data pengujian, yang masing-masing digunakan untuk melatih dan menguji model machine learning.

2. Pemrosesan Data

Sebelum data dapat digunakan untuk melatih model machine learning, dilakukan beberapa langkah preprocessing sebagai berikut:

a. Normalisasi Data: Data numerik dinormalisasi ke dalam rentang 0 hingga 1 menggunakan metode min-max scaling untuk memastikan bahwa semua fitur memiliki kontribusi yang seimbang dalam model.
b. Pembersihan Data: Data yang mengandung nilai yang hilang atau outliers diidentifikasi dan dihapus untuk meningkatkan akurasi model.
c. Pengkodean Fitur Kategorikal: Fitur-fitur kategorikal, seperti jenis protokol (TCP, UDP, ICMP), dikodekan menggunakan metode one-hot encoding untuk memastikan bahwa mereka dapat diolah oleh model machine learning.
d. Pembagian Data: Dataset dibagi menjadi dua bagian: 70% untuk pelatihan dan 30% untuk pengujian, untuk menghindari overfitting dan memastikan generalisasi model.

3. Pemilihan Model Machine Learning

Penelitian ini menggunakan tiga algoritma machine learning yang populer dan efektif untuk deteksi intrusi:

a. Decision Tree: Algoritma ini membagi dataset menjadi subset berdasarkan fitur yang paling informatif, dengan tujuan membentuk model yang dapat digunakan untuk prediksi dengan interpretasi yang mudah.
b. Random Forest: Sebagai metode ensemble, Random Forest menggabungkan hasil dari beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko overfitting.
c. Support Vector Machine (SVM): SVM digunakan untuk memisahkan data dari dua kelas dengan margin maksimum, dan kernel trick diterapkan untuk menangani data yang tidak linear.

4. Pelatihan dan Pengujian Model

Setelah data diproses, masing-masing model machine learning dilatih menggunakan dataset pelatihan. Proses pelatihan melibatkan penyesuaian parameter model untuk meminimalkan error dan meningkatkan akurasi prediksi. Setelah model dilatih, mereka diuji menggunakan dataset pengujian untuk mengevaluasi kinerjanya.

a. Cross-validation: Teknik 10-fold cross-validation digunakan untuk memastikan bahwa model tidak overfit dan mampu melakukan generalisasi pada data yang belum pernah dilihat sebelumnya.

Jurnal Minfo Polgan
Volume 13, Nomor 2, September 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i2.14074 p-ISSN : 2089-9424

b. Hyperparameter Tuning: Proses penyesuaian hyperparameter dilakukan menggunakan grid search untuk menemukan kombinasi parameter yang menghasilkan kinerja terbaik.

HASIL DAN PEMBAHASAN

Setelah dilakukan pelatihan dan pengujian terhadap model Decision Tree, Random Forest, dan Support Vector Machine (SVM), hasil evaluasi kinerja model berdasarkan dataset pengujian ditunjukkan dalam Tabel 1.

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)	AUC
Decision Tree	92.5	91.0	89.5	90.2	0.88
Random Forest	96.8	95.5	94.7	95.1	0.94
Support Vector Machine (SVM)	94.3	93.2	92.8	93.0	0.90

Tabel 1. Kinerja Model Machine Learning pada Dataset Pengujian

Dari Tabel 1, dapat dilihat bahwa algoritma Random Forest menunjukkan kinerja terbaik dengan akurasi 96.8%, presisi 95.5%, recall 94.7%, F1-Score 95.1%, dan nilai AUC 0.94. Sementara itu, Decision Tree dan SVM juga menunjukkan kinerja yang baik, namun sedikit lebih rendah dibandingkan Random Forest.

Pembahasan Hasil

a. Akurasi dan Efektivitas Model. Random Forest menunjukkan akurasi tertinggi di antara ketiga model yang diuji, yang menunjukkan kemampuannya dalam menangani kompleksitas data dan mengurangi risiko overfitting. Kinerja yang lebih tinggi dari Random Forest dapat dikaitkan dengan sifat ensemble-nya, yang menggabungkan prediksi dari banyak pohon keputusan untuk menghasilkan hasil yang lebih stabil dan akurat.

b. Presisi dan Recall. Presisi dan recall yang tinggi pada Random Forest menunjukkan bahwa model ini tidak hanya akurat dalam mendeteksi serangan yang sebenarnya terjadi, tetapi juga efektif dalam meminimalkan false positives. Decision Tree, meskipun akurat, memiliki nilai recall yang sedikit lebih rendah, menunjukkan bahwa model ini cenderung kurang sensitif dalam mendeteksi beberapa jenis serangan yang lebih sulit dikenali.

c. ROC Curve dan AUC. Nilai AUC yang tinggi pada Random Forest (0.94) mengindikasikan bahwa model ini memiliki kemampuan yang sangat baik dalam membedakan antara lalu lintas jaringan normal dan lalu lintas yang mengandung serangan. Ini menunjukkan bahwa Random Forest adalah pilihan yang lebih unggul untuk diterapkan dalam sistem deteksi intrusi yang memerlukan tingkat keandalan tinggi.

Diagram dan Visualisasi.Untuk membantu memahami hasil penelitian, berikut adalah deskripsi diagram yang dapat digunakan:

Confusion Matrix. Confusion Matrix menunjukkan performa masing-masing model dalam klasifikasi intrusi pada dataset pengujian. Diagram ini memperlihatkan jumlah true positives, false positives, true negatives, dan false negatives untuk setiap model.

	Predicted Positive	Predicted Negative
Actual Positive	475	25
Actual Negative	15	485

Tabel 2. Confusion Matrix untuk Random Forest

ROC Curve memperlihatkan trade-off antara true positive rate dan false positive rate untuk ketiga model yang diuji. Semakin dekat kurva ROC ke sudut kiri atas, semakin baik kinerja model.

Bagan 2. ROC Curve . Random Forest: Kurva terletak dekat dengan sudut kiri atas, menunjukkan nilai AUC yang tinggi. Decision Tree: Kurva berada sedikit di bawah Random Forest. SVM: Kurva

Jurnal Minfo Polgan
Volume 13, Nomor 2, September 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i2.14074 p-ISSN : 2089-9424

berada di antara Random Forest dan Decision Tree. Diagram Bar: Akurasi, Presisi, Recall, dan F1-Score. Diagram bar dapat digunakan untuk membandingkan akurasi, presisi, recall, dan F1-Score dari ketiga model.

Model	Akurasi	Presisi	Recall	F1- Score
Decision Tree	92.5	91.0	89.5	90.2
Random Forest	96.8	95.5	94.7	91.5
SVM	94.3	93.2	92.8	93.0

Tabel 3. Perbandingan Kinerja Model Machine Learning

KESIMPULAN

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model machine learning untuk deteksi intrusi pada jaringan komputer, dengan fokus pada tiga algoritma utama: Decision Tree, Random Forest, dan Support Vector Machine (SVM). Berdasarkan hasil penelitian, dapat disimpulkan beberapa poin utama sebagai berikut:

Keefektifan Machine Learning dalam Deteksi Intrusi: Semua model machine learning yang diuji menunjukkan kemampuan yang baik dalam mendeteksi intrusi pada jaringan komputer. Namun, model Random Forest secara konsisten memberikan kinerja terbaik dengan akurasi, presisi, recall, dan F1-score yang lebih tinggi dibandingkan dengan Decision Tree dan SVM. Ini menunjukkan bahwa machine learning, khususnya Random Forest, merupakan pendekatan yang sangat efektif untuk mendeteksi serangan siber.

Keunggulan Random Forest: Random Forest menonjol sebagai algoritma terbaik dalam penelitian ini, terutama karena kemampuannya untuk mengatasi kompleksitas data jaringan dan mengurangi risiko overfitting. Dengan tingkat akurasi mencapai 96.8% dan nilai AUC 0.94, Random Forest terbukti lebih handal dalam membedakan antara lalu lintas jaringan normal dan anomali yang menandakan adanya intrusi.

Implikasi dan Kontribusi Penelitian: Hasil penelitian ini memberikan kontribusi signifikan dalam bidang keamanan jaringan, dengan menunjukkan bahwa implementasi machine learning, terutama dengan algoritma Random Forest, dapat meningkatkan kemampuan deteksi intrusi secara signifikan. Temuan ini dapat digunakan sebagai landasan untuk pengembangan lebih lanjut dari sistem deteksi intrusi (IDS) berbasis machine learning yang lebih adaptif dan responsif terhadap ancaman siber.

Tantangan dan Arah Penelitian Selanjutnya: Meskipun penelitian ini menunjukkan hasil yang positif, beberapa tantangan masih perlu diatasi, termasuk kebutuhan untuk mengurangi false positives dan mengembangkan model yang dapat mendeteksi serangan baru yang belum pernah terjadi sebelumnya (zero-day attacks). Penelitian masa depan diharapkan dapat mengeksplorasi integrasi machine learning dengan teknologi lain, seperti blockchain dan edge computing, untuk lebih meningkatkan keamanan jaringan.

REFERENSI

Alotaibi, F. S. (2019). Implementation of Machine Learning Model to Predict Heart Failure Disease. International Journal of Advanced Computer Science and Applications, 10(6), 261–268. https://doi.org/10.14569/IJACSA.2019.0100637

Bhanu, K. N., Jasmine, H. J., & Mahadevaswamy, H. S. (2020). Machine learning Implementation in IoT based Intelligent System for Agriculture. 2020 International Conference for Emerging Technology, INCET 2020. https://doi.org/10.1109/INCET49848.2020.9153978

Bororing, G. M. G. (2024). PENGEMBANGAN ALGORITMA MACHINE LEARNING UNTUK MENDETEKSI ANOMALI DALAM JARINGAN KOMPUTER. Jurnal Review Pendidikan Dan Pengajaran (JRPP), 7(1), 1361–1368. https://doi.org/10.31004/JRPP.V7I1.25176

Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing Machine Learning in Health Care — Addressing Ethical Challenges. New England Journal of Medicine, 378(11), 981–983. https://doi.org/10.1056/NEJMP1714229/SUPPL_FILE/NEJMP1714229_DISCLOSURES.PDF

Heizmann, M., Braun, A., Glitzner, M., Günther, M., Hasna, G., Klüver, C., Krooß, J., Marquardt, E., Overdick, M., & Ulrich, M. (2022). Implementing machine learning: Chances and challenges. At-Automatisierungstechnik, 70(1), 90–101. https://doi.org/10.1515/AUTO-2021-0149/MACHINEREADABLECITATION/RIS

Mall, S., Srivastava, A., Mazumdar, B. D., Mishra, M., Bangare, S. L., & Deepak, A. (2022). Implementation of machine learning techniques for disease diagnosis. Materials Today: Proceedings, 51, 2198–2201. https://doi.org/10.1016/J.MATPR.2021.11.274

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: an applied review. International Journal of Remote Sensing, 39(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Singh, S., Ramkumar, K. R., & Kukkar, A. (2021). Machine Learning Techniques and Implementation of Different ML Algorithms. 2021 2nd Global Conference for Advancement in Technology, GCAT 2021. https://doi.org/10.1109/GCAT52182.2021.9586806

Xie, Y., Ebad Sichani, M., Padgett, J. E., & DesRoches, R. (2020). The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. Https://Doi.Org/10.1177/8755293020919419, 36(4), 1769–1801. https://doi.org/10.1177/8755293020919419