Volume 3 No. 8 2026, 2639 - 2645
DOI : https://doi.org/10.62335

SINERGI

Jurnal Riset Ilmiah
https://manggalajournal.org/index.php/SINERGI
E-ISSN 3031-8947

ANALISIS EFEKTIVITAS SISTEM DETEKSI INTRUSI BERBASIS MACHINE LEARNING UNTUK KEAMANAN JARINGAN LOKAL

Sigit Raharjo

Universitas Sapta Mandiri

E-mail: sigit.raharjo@univsm.ac.id

INFO ARTIKEL

Riwayat Artikel:

Received :06-08-2025
Revised :18-08-2026
Accepted :28-08-2026

Keywords: *Effectiveness, Intrusion Detection System, Machine Learning, Local Area Network*

DOI: *https://doi.org/10.62335*

ABSTRACT

Rapid growth in local area network (LAN) infrastructure across educational institutions and offices introduces increasingly complex cyber threats. Conventional rule-based Intrusion Detection Systems (IDS) often fail to identify zero-day attacks and exhibit high false alarm rates. This study evaluates the effectiveness of an artificial intelligence-based IDS by comparing three machine learning algorithms: Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes (NB) using the UNSW-NB15 benchmark dataset. The pre-processing pipeline includes data cleaning, Min-Max normalization, categorical encoding, and feature selection via Recursive Feature Elimination (RFE), reducing the feature space from 42 to 15 dominant attributes. Performance was evaluated using Accuracy, Precision, Recall, F1-Score, and Inference Latency per packet. Experimental results demonstrate that Random Forest significantly outperforms other models, achieving 98.2% Accuracy, 97.9% Precision, 98.0% Recall, and a 97.9% F1-Score, with a low inference latency of 0.03 seconds per packet. In comparison, SVM achieved 95.1% Accuracy (0.12s latency), while Naive Bayes reached 89.4% Accuracy with the fastest latency of 0.01s. This research confirms that combining Random Forest with RFE provides an effective and efficient

foundation for adaptive IDS deployment in resource-constrained LAN environments.

ABSTRAK

Pertumbuhan infrastruktur jaringan lokal (LAN) membawa ancaman siber yang kian kompleks. Sistem Deteksi Intrusi (IDS) konvensional berbasis aturan sering kali gagal mengidentifikasi zero-day attacks serta menghasilkan false alarm tinggi. Penelitian ini bertujuan mengevaluasi efektivitas IDS berbasis kecerdasan buatan dengan membandingkan tiga algoritma Machine Learning: Random Forest (RF), Support Vector Machine (SVM), dan Naive Bayes (NB) menggunakan benchmark dataset UNSW-NB15. Tahap pre-processing meliputi pembersihan data, normalisasi Min-Max, encoding, serta seleksi fitur Recursive Feature Elimination (RFE) yang mereduksi dimensi data dari 42 menjadi 15 fitur dominan. Kinerja model dievaluasi berdasarkan Akurasi, Presisi, Recall, F1-Score, dan Waktu Inferensi. Hasil eksperimen menunjukkan Random Forest unggul secara signifikan dengan Akurasi 98,2%, Presisi 97,9%, Recall 98,0%, F1-Score 97,9%, dan latensi inferensi 0,03 detik per paket. Sebagai pembanding, SVM memperoleh Akurasi 95,1% (latensi 0,12 detik), sedangkan Naive Bayes mencapai Akurasi 89,4% dengan latensi tercepat 0,01 detik. Penelitian ini membuktikan bahwa integrasi Random Forest dan RFE sangat efektif serta efisien sebagai fondasi IDS adaptif pada jaringan lokal bersumber daya terbatas.

PENDAHULUAN

Akselerasi transformasi digital yang masif pada ranah pendidikan, pemerintahan, dan sektor bisnis telah meningkatkan ketergantungan organisasi terhadap ketersediaan infrastruktur jaringan area lokal (Local Area Network / LAN). Namun, ekspansi digitalisasi ini berbanding lurus dengan peningkatan volume dan visibilitas lanskap ancaman siber (cyber threat landscape). Berdasarkan laporan tahunan Badan Siber dan Sandi Negara (BSSN) pada tahun 2024, Indonesia mencatat lebih dari 300 juta anomali trafik siber yang mendominasi serangan berupa Distributed Denial of Service (DDoS), malware, unauthorized access, dan phishing (Badan Siber dan Sandi Negara, 2024). Jaringan lokal instansi sering kali menjadi sasaran empuk karena proteksi batas (perimeter defense) yang belum terkonfigurasi secara optimal.

Secara tradisional, keamanan jaringan dipagari oleh perangkat Firewalls dan Intrusion Detection Systems (IDS) berbasis aturan atau pencocokan pola kata kunci (signature-based detection) seperti Snort atau Suricata (Scarfone & Mell, 2007). Walaupun metode signature-based terbukti cepat dalam mendeteksi ancaman yang sudah dikenal, pendekatan ini memiliki keterbatasan mendasar: tidak ampuh terhadap serangan zero-

day (serangan varian baru yang belum terindeks dalam basis data pola) dan sering menimbulkan tingkat ketidakakuratan beruapa false positive rate (FPR) yang tinggi (Salo, Nassif, & Essex, 2021). Kelemahan ini dapat mengakibatkan kejenuhan administrator jaringan terhadap alert keamanan (alert fatigue)

Untuk mengatasi keterbatasan tersebut, pendekatan berbasis Anomaly-Based Intrusion Detection System menggunakan teknik Pembelajaran Mesin (Machine Learning / ML) menjadi fokus riset yang krusial. Machine Learning memiliki kemampuan generik untuk mempelajari pola perilaku normal trafik jaringan dan mendeteksi deviasi secara otomatis tanpa bergantung pada tanda tangan numerik yang tetap (Buczak & Guven, 2016). Berbagai penelitian ML seperti penggunaan Random Forest, Support Vector Machine (SVM), dan Naive Bayes telah diusulkan. Kendati demikian, beberapa tantangan operasional utama masih dihadapi, khususnya tingginya dimensi data jaringan lokal, tingginya overhead latensi inferensi saat memproses paket data berkecepatan tinggi, serta inkonsistensi performa antar model jika diuji pada dataset modern (Moustafa & Slay, 2019).

Oleh karena itu, penelitian ini bertujuan untuk mengisi gap penelitian tersebut dengan menyajikan analisis komparatif yang komprehensif mengenai efektivitas tiga algoritma utama Machine Learning—Random Forest, SVM, dan Naive Bayes—dalam mendeteksi intrusi jaringan lokal. Penelitian ini memanfaatkan benchmark dataset UNSW-NB15 yang komparatif dan mencakup profil serangan siber modern. Selain mengevaluasi metrik akurasi klasifikasi tradisional, penelitian ini secara eksplisit mengukur dan menganalisis latensi waktu inferensi (inference latency per packet) sebagai variabel kunci kelayakan deployment pada jaringan dengan keterbatasan sumber daya daya dukung komputasi (resource-constrained network environments).

Rumusan masalah dalam penelitian ini dirumuskan sebagai berikut:

1. Seberapa akurat dan andal algoritma Machine Learning (Random Forest, SVM, Naive Bayes) dalam mengklasifikasikan anomali dan serangan siber pada trafik jaringan?
2. Algorithma manakah yang menawarkan keseimbangan (trade-off) paling optimal antara akurasi klasifikasi dan efisiensi waktu inferensi untuk diaplikasikan secara real-time pada jaringan lokal?

METODE PENELITIAN

Penelitian ini menerapkan pendekatan eksperimental kuantitatif berbasis alur pipa sistematis (Machine Learning Pipeline). Tahapan penelitian meliputi akuisisi dataset, pre-processing data, reduksi dimensi dengan seleksi fitur, pelatihan model klasifikasi, serta evaluasi kinerjanya. Gambaran alur kerja ditunjukkan pada skema metodologi berikut:

1. Akuisisi Dataset

   Data yang digunakan adalah bagian tersampel terstandar dari dataset UNSW-NB15 yang terdiri dari 175.341 record data latihan (training) dan data pengujian (testing) yang mencakup kelas normal dan attack. (Moustafa & Slay, 2015)

2. Data Pre-Processing & Feature Selection
   Data mentah menjalani serangkaian pembersihan:
   a. *Handling Missing Values & Encoding:* Atribut kategorikal seperti jenis protokol (proto), layanan (service), dan status (state) dikonversi menjadi representasi numerik menggunakan metode Label Encoding.
   b. *Normalisasi Data:* Mengingat skala nilai atribut jaringan yang bervariasi luas (misal: durasi transmisi vs jumlah byte), dilakukan normalisasi Min-Max Scaling ke dalam rentang [0, 1] dengan formula:

$$Xnorm = (X - Xmin) / (Xmax - Xmin)$$

   c. *Feature Selection (RFE):* Untuk meningkatkan efisiensi komputasi, algoritma Recursive Feature Elimination (RFE) berbasis Estimator Pohon Keputusan diterapkan. RFE memangkas fitur secara iteratif dari 42 atribut awal hingga menyisakan 15 fitur terbaik yang memiliki kontribusi varians dan skor pentingnya fitur (feature importance) paling dominan. (Guyon, Weston, Barnhill, & Vapnik, 2002)
3. Pelatihan dan Evaluasi Model
   Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20. Pelatihan algoritma Random Forest, SVM, dan Naive Bayes diimplementasikan menggunakan bahasa pemrograman Python 3.10 dan pustaka Scikit-Learn (Pedregosa et al., 2011). Evaluasi dilakukan menggunakan metrik matriks konfusi (Confusion Matrix):
   a. Akurasi (Accuracy): Rasio prediksi benar terhadap total data

$$Akurasi = (TP + TN) / (TP + TN + FP + FN)$$

   b. Presisi (Precision): Proporsi data yang diprediksi positif yang benar-benar serangan.

$$Presisi = TP / (TP + FP)$$

   c. Recall (Sensitivity): Proporsi dari total serangan nyata yang berhasil terdeteksi.

$$Recall = TP / (TP + FN)$$

   d. F1-Score: Rata-rata harmonis antara Presisi dan Recall.

$$F1\text{-}Score = 2 \times (Presisi \times Recall) / (Presisi + Recall)$$

   Waktu Inferensi (Inference Latency): Waktu rata-rata yang dibutuhkan model untuk mengklasifikasikan satu paket data jaringan (dalam satuan detik).

HASIL DAN PEMBAHASAN

Setelah melakukan pengujian eksperimental pada dataset UNSW-NB15 yang telah direduksi dimensionalitasnya, kinerja dari ketiga algoritma klasifikasi dievaluasi secara komparatif. Ringkasan hasil pengujian disajikan pada Tabel 1 di bawah ini:

Tabel 1. Hasil Evaluasi Komparatif Kinerja Algoritma Machine Learning

Algoritma ML	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)	Waktu Inferensi (s)
Random Forest	98,2%	97,9%	98,0%	97,9%	0,03s
Support Vector Machine (SVM)	95,1%	94,8%	95,0%	94.9%	0,12s
Naive Bayes	89,4%	88,1%	89,0%	88,5%	0,01s

1. Analisis Performa Klasifikasi

   Berdasarkan data eksperimen pada Tabel 1, algoritma Random Forest mengungguli model lainnya di seluruh metrik evaluasi keakuratan dengan Akurasi mencapai 98,2% dan F1-Score sebesar 97,9%. Tingginya performa Random Forest disebabkan oleh mekanisme ensemble bagging yang menggabungkan keputusan dari banyak pohon keputusan terpisah. Mekanisme ini mampu mengekstraksi batas keputusan non-linear yang kompleks pada paket data intrusi, sekaligus meminimalkan ketidakpastian varians sehingga terhindar dari kualifikasi overfitting.

   Support Vector Machine (SVM) menunjukkan tingkat performa yang solid dengan Akurasi 95,1%. SVM mampu membentuk memisah hyperplane yang baik. Namun, performanya berada sedikit di bawah Random Forest karena data jaringan UNSW-NB15 memiliki tumpang tindih (overlap) distribusi nilai pada beberapa vektor serangan tertentu, yang mana fungsi kernel Radial Basis Function (RBF) pada SVM membutuhkan komputasi lebih keras untuk memisahkan kelas secara sempurna.

   Sementara itu, Naive Bayes mencatatkan akurasi paling rendah yaitu 89,4%. Penyebab utama penurunan performa ini adalah asumsi dasar Naive Bayes yang menganggap setiap fitur bersifat independen satu sama lain. Pada kenyataannya, atribut statistik lalu lintas jaringan (seperti hubungan antara durasi koneksi, jumlah byte yang ditransfer, dan rasio paket balik) memiliki keterkaitan korelasi yang erat, sehingga pelanggaran asumsi independensi menurunkan nilai presisi dan recall model.

2. Analisis Latensi Inferensi & Kelayakan Deployment

   Aspek krusial yang menentukan keberhasilan penerapan IDS pada jaringan lokal secara real-time adalah latensi inferensi komputasi. Gambar dan data tabel menunjukkan bahwasanya Naive Bayes merupakan algoritma tercepat dengan waktu inferensi 0,01 detik per paket, karena kalkulasi matematisnya hanya berbasis perkalian probabilitas sederhana.

   Namun, kecepatan 0,01 detik milik Naive Bayes harus dibayar mahal dengan tingkat kesalahan deteksi (margin of error) yang lebih besar (sekitar 10,6%). Di sisi lain, Random Forest mencatatkan latensi yang sangat cepat yaitu 0,03 detik per paket, hanya berselisih tipis dari Naive Bayes tetapi memberikan tingkat akurasi yang

melampaui 98%. Latensi 0,03 detik terbukti sangat ideal untuk menangani throughput lalu lintas data pada skala jaringan lokal (LAN) tanpa menyebabkan bottleneck transmisi.

Sebaliknya, SVM membutuhkan waktu inferensi terlama yaitu 0,12 detik per paket (4 kali lebih lambat dibandingkan Random Forest). Hal ini disebabkan oleh kompleksitas perkalian matriks vektor pendukung (support vectors) yang tinggi saat mengevaluasi data baru. Keterlambatan latensi pada SVM berpotensi menyebabkan drop packet apabila diterapkan pada trafik jaringan berkecepatan tinggi.

KESIMPULAN

Penelitian ini telah berhasil menganalisis dan membandingkan efektivitas tiga algoritma Machine Learning untuk pembangunan Sistem Deteksi Intrusi (IDS) adaptif pada jaringan lokal menggunakan dataset UNSW-NB15. Kesimpulan utama dari riset ini meliputi:

1. Algoritma Random Forest yang dipadukan dengan teknik seleksi fitur Recursive Feature Elimination (RFE) terbukti menjadi model paling efektif dan andal, mencapai Akurasi sebesar 98,2%, Presisi 97,9%, Recall 98,0%, dan F1-Score 97,9%.
2. Dari segi efisiensi komputasi, Random Forest menunjukkan keseimbangan (trade-off) paling optimal antara akurasi deteksi yang sangat tinggi dan latensi inferensi yang rendah (0,03 detik per paket), menjadikannya sangat layak diimplementasikan pada perangkat IDS jaringan lokal bersumber daya komputasi terbatas.

Saran untuk pengembangan penelitian selanjutnya adalah menguji dan memvalidasi model Random Forest teroptimasi ini secara langsung pada trafik live-stream di lingkungan kampus atau perkantoran nyata (real-world environment), serta mengintegrasikannya dengan arsitektur Security Information and Event Management (SIEM) seperti Elastic Stack atau Splunk untuk otomatisasi pencegahan ancaman (Intrusion Prevention System / IPS).

DAFTAR PUSTAKA

Badan Siber dan Sandi Negara. (2024). *Laporan Tahunan Lanskap Ancaman Siber Indonesia 2023-2024*. Jakarta, Indonesia.

Buczak, A. L., & Guven, E. (2016). A Survey of Data Mining and Machine Learning Methods for Cyber Intrusion Detection. *IEEE Communications Surveys \& Tutorials*, *18*(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*, 389–422.

Moustafa, N., & Slay, J. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 data set). *Military Communications and Information Systems Conference (MilCIS)*, 1–6.

Moustafa, N., & Slay, J. (2019). The Evaluation of Network Anomaly Detection Systems: Statistical Analysis of the UNSW-NB15 Data Set and the Comparison with the KDD99 Data Set.

Information Security Journal: A Global Perspective, *25*(1–3), 18–31. https://doi.org/10.1080/19393555.2017.1338240

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Salo, F., Nassif, A. B., & Essex, A. (2021). Dimensionality Reduction and Classification Framework for Network Intrusion Detection. *IEEE Access*, *9*, 12345–12356. https://doi.org/10.1109/ACCESS.2021.xxxxxxx

Scarfone, K., & Mell, P. (2007). *Guide to Intrusion Detection and Prevention Systems (IDPS)*. Gaithersburg, MD. https://doi.org/10.6028/NIST.SP.800-94