Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 5, Mei 2025, Hal. 1365-1376
DOI: https://doi.org/10.52436/1.jpti.790
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Deteksi Anomali Trafik Jaringan dan Aktivitas Pengguna Menggunakan Isolation Forest untuk Meningkatkan Keamanan Jaringan

Khabib Adi Nugroho[*1], Taqwa Hariguna[2], Azhari Shouni Barkah[3]

[1,2]Magister Ilmu Komputer, Universitas Amikom Purwokerto
[3]Informatika, Universitas Amikom Purwokerto
Email: [1]23ma41d018@students.amikompurwokerto.ac.id, [2]taqwa@amikompurwokerto.ac.id, [3]azhari@amikompurwokerto.ac.id

Abstrak

Peningkatan kompleksitas serangan siber menuntut pengembangan metode deteksi anomali yang lebih adaptif dan efisien. Sistem deteksi intrusi berbasis tanda tangan (signature-based IDS) memiliki keterbatasan dalam mengenali serangan baru atau serangan zero-day, sehingga dibutuhkan pendekatan berbasis pembelajaran mesin untuk mengidentifikasi anomali tanpa bergantung pada pola serangan yang telah terdokumentasi sebelumnya. Tujuan dari penelitian ini adalah untuk mengembangkan sistem deteksi intrusi menggunakan algoritma Isolation Forest yang efektif dalam mendeteksi anomali dalam lalu lintas jaringan. Penelitian ini mengimplementasikan Isolation Forest untuk menganalisis dataset lalu lintas jaringan CICIDS 2017, yang sebelumnya diproses melalui langkah-langkah preprocessing, termasuk pemilihan fitur, normalisasi data, dan penanganan nilai yang hilang. Model ini dilatih menggunakan hanya data normal traffic untuk membangun baseline perilaku jaringan, yang kemudian digunakan untuk mendeteksi anomali pada seluruh dataset. Evaluasi dilakukan dengan metrik akurasi, presisi, recall, F1-score, dan False Positive Rate (FPR), yang menunjukkan hasil yang menggembirakan. Model mencapai akurasi 86,50%, dengan presisi 83,86%, yang mengindikasikan bahwa sebagian besar prediksi anomali adalah benar. Pengaturan ambang batas deteksi anomali pada persentil 5% menghasilkan FPR yang rendah 0,97%, yang berperan penting dalam mengurangi alarm palsu dan meningkatkan efisiensi analisis keamanan. Penelitian ini menunjukkan bahwa Isolation Forest efektif dalam mendeteksi anomali dalam lalu lintas jaringan, dengan tingkat false positives yang rendah, menjadikannya solusi yang menjanjikan dalam meningkatkan sistem deteksi intrusi berbasis perilaku. Dampak dari penelitian ini memberikan kontribusi signifikan dalam pengembangan sistem deteksi intrusi berbasis pembelajaran mesin, yang dapat lebih responsif terhadap ancaman siber yang terus berkembang.

Kata kunci: *Deteksi Anomali, Intrusion Detection System (IDS), Isolation Forest, Keamanan Siber, Machine Learning.*

Network Traffic and User Activity Anomaly Detection Using Isolation Forest to Improve Network Security

Abstract

The increasing complexity of cyber-attacks demands the development of more adaptive and efficient anomaly detection methods. Signature-based intrusion detection systems (IDS) have limitations in recognizing new attacks or zero-day attacks, so a machine learning-based approach is needed to identify anomalies without relying on previously documented attack patterns. The purpose of this research is to develop an intrusion detection system using the Isolation Forest algorithm that is effective in detecting anomalies in network traffic. This research implements Isolation Forest to analyze the 2017 CICIDS network traffic dataset, which was previously processed through preprocessing steps, including feature selection, data normalization, and missing value handling. The model was trained using only normal traffic data to establish a baseline of network behavior, which was then used to detect anomalies in the entire dataset. Evaluation was conducted using accuracy, precision, recall, F1-score, and False Positive Rate (FPR) metrics, which showed encouraging results. The model achieved an accuracy of 86.50%, with a precision of 83.86%, indicating that most of the anomaly predictions were correct. Setting the anomaly detection threshold at the 5% percentile results in a low FPR of 0.97%, which plays an important role in reducing false alarms and improving the efficiency of security analysis. This research shows that Isolation Forest is effective in detecting anomalies in network traffic, with a low rate of false positives, making it a promising solution in improving behavior-based intrusion detection systems. The impact of this research makes a significant

p-ISSN: 2775-4227
e-ISSN: 2775-4219

contribution to the development of machine learning-based intrusion detection systems, which can be more responsive to evolving cyber threats.

Keywords: *Anomaly Detection, Cyber Security, Intrusion Detection System (IDS), Isolation Forest, Machine Learning.*

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah membawa manfaat besar bagi berbagai sektor, namun juga membuka celah bagi peningkatan jumlah dan kompleksitas serangan siber. Serangan siber modern semakin canggih dengan memanfaatkan teknik seperti malware polimorfik, serangan terdistribusi, serta eksploitasi celah keamanan yang belum diketahui sebelumnya [1], [2]. Keamanan jaringan konvensional yang bergantung pada firewall dan sistem deteksi intrusi berbasis tanda tangan semakin kurang efektif dalam menghadapi ancaman yang berkembang ini [2], [3]. Selain itu, pertumbuhan layanan berbasis cloud dan Internet of Things (IoT) memperbesar permukaan serangan, sehingga meningkatkan kebutuhan akan sistem keamanan yang lebih adaptif dan cerdas dalam mendeteksi ancaman siber secara real-time.

Pendekatan deteksi intrusi tradisional, seperti sistem deteksi berbasis tanda tangan dan aturan, memiliki keterbatasan dalam menghadapi serangan yang belum terdokumentasi [4]. Sistem berbasis tanda tangan hanya dapat mengenali pola serangan yang telah diketahui, sehingga tidak mampu mendeteksi ancaman zero-day yang memanfaatkan celah keamanan baru [5]. Sementara itu, pendekatan berbasis aturan memerlukan pembaruan manual yang berkelanjutan, yang tidak hanya memakan waktu tetapi juga berisiko menimbulkan false positives yang tinggi [6]. Untuk mengatasi keterbatasan ini, deteksi berbasis anomali dengan algoritma pembelajaran mesin telah menjadi solusi yang menjanjikan karena mampu mengenali aktivitas mencurigakan berdasarkan pola perilaku yang tidak biasa, bukan hanya berdasarkan pola serangan yang telah dikenal sebelumnya.

Sistem deteksi intrusi berbasis tanda tangan memiliki kelemahan dalam mendeteksi serangan baru yang belum terdokumentasi. Ketergantungan pada database tanda tangan menyebabkan keterlambatan dalam mengenali ancaman baru, sehingga memberi peluang bagi penyerang untuk mengeksploitasi kelemahan yang belum dikenali [7], [8]. Selain itu, sistem berbasis aturan yang menggunakan parameter yang didefinisikan secara manual sulit untuk diperbarui secara dinamis dan sering kali menghasilkan tingkat false positive yang tinggi [9]. Akibatnya, banyak aktivitas jaringan yang sebenarnya tidak berbahaya salah diklasifikasikan sebagai ancaman, yang pada akhirnya membebani tim keamanan siber dengan peringatan yang tidak relevan [10], [11].

Tantangan lain dalam deteksi intrusi adalah meningkatnya volume lalu lintas jaringan yang harus dipantau dan dianalisis secara efisien. Sistem konvensional mengalami kesulitan dalam menangani jumlah data yang besar dan kompleks, terutama dengan adanya enkripsi yang dapat menyembunyikan pola komunikasi berbahaya [12]. Selain itu, teknik pengelakan serangan seperti tunneling dan obfuscation semakin sering digunakan oleh penyerang untuk menghindari deteksi oleh sistem keamanan tradisional [13], [14]. Sehingga, diperlukan pendekatan deteksi yang lebih adaptif dan berbasis data untuk meningkatkan ketahanan terhadap ancaman yang terus berkembang.

Penelitian ini bertujuan untuk menerapkan model deteksi anomali berbasis algoritma Isolation Forest guna meningkatkan kemampuan deteksi aktivitas mencurigakan dalam lalu lintas jaringan dan perilaku pengguna. Berbeda dengan metode berbasis tanda tangan yang bergantung pada data historis serangan, Isolation Forest menggunakan pendekatan unsupervised learning untuk mengidentifikasi penyimpangan dari pola perilaku normal tanpa memerlukan label serangan sebelumnya [15]. Pendekatan ini memungkinkan deteksi ancaman baru, termasuk serangan zero-day dan metode intrusi canggih lainnya, dengan menganalisis karakteristik lalu lintas jaringan dan perilaku pengguna.

Selain mengimplementasikan model Isolation Forest, penelitian ini juga mengevaluasi kinerjanya dibandingkan dengan metode deteksi tradisional [16]. Untuk mendapatkan analisis yang komprehensif, evaluasi dilakukan dengan menggunakan metrik utama seperti akurasi, presisi, recall, dan tingkat false positive. Penelitian ini juga menyelidiki kelayakan penerapan sistem berbasis Isolation Forest dalam lingkungan keamanan siber secara real-time guna memastikan skalabilitas dan kemampuannya dalam mendeteksi ancaman siber secara adaptif.

Penelitian ini menggunakan pendekatan berbasis machine learning untuk mendeteksi anomali dalam lalu lintas jaringan. Data yang digunakan berasal dari dataset CICIDS 2017, yang mencakup berbagai jenis serangan siber serta lalu lintas normal. Langkah awal penelitian meliputi preprocessing data, termasuk normalisasi fitur numerik, encoding fitur kategorikal, serta pembuatan jendela waktu untuk membentuk data sekuensial yang dapat digunakan oleh model deteksi.

Selanjutnya, model Isolation Forest dilatih menggunakan data lalu lintas normal untuk membangun baseline pola aktivitas jaringan yang dianggap aman. Setelah itu, seluruh data diuji untuk mengidentifikasi anomali berdasarkan skor isolasi yang dihasilkan oleh model. Penentuan ambang batas dilakukan dengan

p-ISSN: 2775-4227
e-ISSN: 2775-4219

mempertimbangkan distribusi skor anomali guna mengurangi tingkat false positive. Akhirnya, evaluasi kinerja model dilakukan dengan membandingkan hasil deteksi anomali dengan label serangan aktual dalam dataset. Kesimpulan penelitian ini akan digunakan sebagai dasar untuk pengembangan sistem keamanan berbasis machine learning yang lebih adaptif dan efisien di masa depan [17].

Dalam deteksi anomali pada sistem jaringan, beberapa metode telah banyak digunakan, seperti One-Class Support Vector Machine (SVM), Autoencoders, Random Forest, dan Isolation Forest. One-Class SVM efektif dalam memisahkan data anomali dari data normal, namun kesulitan menangani data berdimensi tinggi dan memerlukan parameter yang tepat. Autoencoders, yang menggunakan jaringan neural untuk mendeteksi anomali berdasarkan rekonstruksi error, efektif namun rentan terhadap overfitting jika data tidak cukup [18]. Random Forest dapat menangani dataset besar dan kompleks dengan baik, namun dapat kesulitan mendeteksi anomali yang lebih halus [19]. Di antara metode ini, Isolation Forest menonjol karena kemampuannya dalam menangani data besar dan berdimensi tinggi secara efisien tanpa bergantung pada data berlabel, menjadikannya pilihan unggul untuk deteksi anomali. Namun, keefektifannya bergantung pada pemilihan parameter seperti jumlah pohon dan contamination rate.

Tujuan dari penelitian ini adalah untuk mengeksplorasi efektivitas Isolation Forest dalam mendeteksi anomali dalam lalu lintas jaringan dan untuk membandingkannya dengan metode lain seperti One-Class SVM, Autoencoders, dan Random Forest. Dengan melakukan perbandingan ini, penelitian bertujuan untuk memberikan pemahaman yang lebih dalam tentang kelebihan dan kelemahan masing-masing metode serta untuk menunjukkan apakah Isolation Forest dapat menjadi solusi yang lebih efisien dan tepat dalam konteks deteksi anomali jaringan. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan sistem deteksi intrusi berbasis pembelajaran mesin yang lebih efektif, dengan fokus pada deteksi serangan yang lebih adaptif dan skalabel pada jaringan yang semakin kompleks.

2. METODE PENELITIAN

2.1 Deskripsi Dataset

Penelitian ini menggunakan dataset CICIDS 2017, yang merupakan kumpulan data lalu lintas jaringan yang mencerminkan aktivitas normal dan berbagai jenis serangan siber. Dataset ini terdiri dari lebih dari dua juta entri dengan 53 fitur, yang mencakup informasi penting seperti durasi aliran data, jumlah paket yang dikirim, serta statistik interval antar paket. Setiap sampel dalam dataset dikategorikan sebagai lalu lintas normal atau sebagai salah satu dari berbagai jenis serangan siber, termasuk serangan Denial of Service (DoS), Distributed Denial of Service (DDoS), pemindaian port, serta serangan berbasis web.

Untuk tujuan penelitian ini, fitur kategorikal seperti Destination Port dan Attack Type dihapus agar tidak mempengaruhi model pembelajaran mesin. Hanya fitur numerik yang dipilih untuk dianalisis, karena fitur ini memberikan informasi yang relevan dalam menganalisis lalu lintas jaringan dan mengidentifikasi anomali. Pemilihan fitur didasarkan pada pertimbangan relevansi masing-masing fitur terhadap karakteristik serangan jaringan, yang mencakup statistik lalu lintas seperti durasi aliran data, jumlah paket, serta berbagai flag paket data yang seringkali berperan dalam membedakan antara aktivitas normal dan aktivitas yang mencurigakan.

Proses pemilihan fitur ini penting karena fitur-fitur yang tidak relevan atau kurang memberikan informasi akan memperburuk kinerja model. Dengan memilih fitur yang secara langsung berhubungan dengan aktivitas jaringan dan jenis serangan yang mungkin terjadi, penelitian ini bertujuan untuk meningkatkan akurasi deteksi anomali dalam lalu lintas jaringan. Fitur-fitur yang dipilih membantu model untuk mengenali pola-pola yang menunjukkan serangan atau anomali lainnya secara lebih efisien.

2.2 Preprocessing Data

Langkah pertama dalam preprocessing data adalah menangani nilai yang hilang atau missing values. Data yang tidak lengkap dapat mengurangi kualitas model, Sehingga, baris yang mengandung nilai yang hilang dihapus dari dataset. Hal ini bertujuan untuk memastikan bahwa model hanya bekerja dengan data yang lengkap dan terstruktur, sehingga proses pelatihan dan evaluasi dapat berjalan dengan lancar tanpa gangguan yang disebabkan oleh nilai yang hilang. Penghapusan baris dengan nilai yang hilang juga meminimalkan risiko adanya ketidakakuratan dalam hasil prediksi model yang dapat terjadi jika nilai yang hilang diproses secara salah atau diisi dengan cara yang tidak tepat.

Setelah menangani nilai yang hilang, langkah selanjutnya adalah normalisasi fitur numerik untuk memastikan bahwa semua fitur dalam dataset berada pada skala yang seragam. Normalisasi dilakukan menggunakan StandardScaler, yang mengubah data agar memiliki rata-rata nol dan standar deviasi satu. Proses normalisasi ini sangat penting karena algoritma pembelajaran mesin, termasuk Isolation Forest, bekerja lebih baik ketika data yang diberikan berada pada skala yang konsisten. Normalisasi dilakukan dengan rumus:

p-ISSN: 2775-4227
e-ISSN: 2775-4219

$$X' = \frac{X-\mu}{\sigma} \qquad (1)$$

di mana X adalah nilai asli dari fitur, μ adalah rata-rata, dan σ adalah standar deviasi dari fitur tersebut. Dengan normalisasi ini, fitur yang memiliki rentang nilai yang sangat besar atau kecil tidak akan mendominasi proses pelatihan model, yang berisiko menurunkan performa model.

Setelah proses normalisasi selesai, dataset yang telah diproses disimpan dalam file cicids_2017_preprocessed.csv. File ini siap digunakan dalam proses pelatihan model Isolation Forest untuk deteksi anomali. Dengan langkah-langkah preprocessing ini, data telah dipersiapkan secara optimal untuk membangun model yang dapat secara efektif mendeteksi anomali dalam lalu lintas jaringan. Proses preprocessing yang baik akan meningkatkan akurasi dan kinerja model dalam mendeteksi serangan atau anomali yang ada dalam data.

2.3 Implementasi Model Isolation Forest

Isolation Forest adalah algoritma pembelajaran mesin berbasis ensemble learning yang dirancang untuk mendeteksi anomali dalam dataset berdimensi tinggi. Algoritma ini bekerja berdasarkan prinsip isolasi, di mana anomali memiliki karakteristik unik yang membuatnya lebih mudah diisolasi dibandingkan data normal. Tidak seperti metode deteksi anomali berbasis kepadatan atau jarak seperti k-nearest neighbors (k-NN) atau Gaussian Mixture Models (GMMs), Isolation Forest tidak bergantung pada distribusi data dan lebih efisien dalam menangani dataset besar dengan kompleksitas waktu linier $O(n \log n)$ [20].

Isolation Forest membangun sejumlah pohon isolasi (Isolation Trees), yang masing-masing dibuat dengan membagi data secara acak berdasarkan fitur yang dipilih secara acak. Ketika sebuah titik data diisolasi lebih cepat dibandingkan yang lain, kemungkinan besar titik tersebut adalah anomali. Proses ini menghasilkan panjang jalur rata-rata dari akar ke daun, di mana data dengan jalur yang lebih pendek dianggap lebih anomali dibandingkan dengan data yang memiliki jalur lebih panjang [21], [22].

Secara matematis, skor anomali s(x, n) untuk sebuah sampel data x dalam dataset dengan n sampel dihitung dengan rumus:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \qquad (2)$$

di mana E(h(x)) adalah panjang jalur rata-rata dari pohon isolasi dan c(n) adalah nilai ekspektasi panjang jalur untuk data normal dalam dataset dengan n sampel. Nilai s(x,n) yang lebih tinggi menunjukkan kemungkinan besar bahwa x adalah anomali.

Pada penelitian ini, model Isolation Forest diterapkan dengan beberapa parameter utama untuk mendeteksi anomali dalam dataset lalu lintas jaringan. Parameter pertama adalah jumlah pohon dalam ensemble model, yang diatur dengan nilai n_estimators = 100. Semakin banyak pohon yang digunakan, semakin baik estimasi model dalam menghitung panjang jalur rata-rata untuk setiap sampel data. Hal ini meningkatkan kemampuan model untuk memisahkan anomali dengan lebih akurat, karena setiap pohon berfungsi untuk mengisolasi sampel dengan cara yang berbeda. Namun, perlu dicatat bahwa jumlah pohon yang sangat besar dapat menyebabkan peningkatan waktu komputasi tanpa memberikan peningkatan signifikan dalam kinerja model.

Selain itu, tingkat kontaminasi (contamination = 0.05) ditentukan untuk mencerminkan persentase sampel dalam dataset yang diperkirakan sebagai anomali. Nilai ini diatur berdasarkan analisis distribusi anomali dalam data yang tersedia. Dalam konteks ini, nilai 0.05 menunjukkan bahwa sekitar 5% dari total data diasumsikan sebagai anomali, yang sesuai dengan proporsi anomali yang biasanya ditemukan dalam banyak dataset keamanan jaringan. Nilai ini sangat penting untuk menyeimbangkan model antara mendeteksi anomali dengan benar dan menghindari terlalu banyak deteksi positif palsu.

Hanya data dengan label Normal Traffic yang digunakan dalam proses pelatihan, karena model ini menggunakan pendekatan unsupervised learning. Hal ini memungkinkan Isolation Forest untuk membangun model dari distribusi data normal tanpa memerlukan label serangan, sehingga lebih adaptif terhadap serangan siber baru yang belum terdokumentasi sebelumnya.

Setelah model dilatih menggunakan data normal, skor anomali dihitung untuk seluruh dataset. Skor ini digunakan untuk menentukan apakah suatu sampel diklasifikasikan sebagai normal atau anomali. Jika skor anomali lebih rendah dari ambang batas yang telah ditetapkan, maka sampel tersebut dikategorikan sebagai anomali.

Setelah skor anomali dihitung, ambang batas deteksi anomali ditentukan berdasarkan persentil ke-5 dari distribusi skor anomali. Pendekatan ini memastikan bahwa hanya 5% dari total data dengan skor anomali terendah yang diklasifikasikan sebagai serangan, sesuai dengan nilai parameter contamination yang telah ditentukan. Pendekatan ini secara matematis dinyatakan sebagai:

p-ISSN: 2775-4227
e-ISSN: 2775-4219

$$Threshold = P_5(S) \quad (3)$$

di mana $P_5(S)$ adalah nilai persentil ke-5 dari distribusi skor anomali S. Jika nilai skor anomali suatu sampel x lebih kecil dari ambang batas ini, maka:

$$\text{Label}(x) = 1 \quad \text{(Anomali)} \quad (4)$$

dan jika lebih besar dari ambang batas:

$$\text{Label}(x) = 0 \quad \text{(Normal)} \quad (5)$$

Metode ini memberikan fleksibilitas dalam menentukan ambang batas berdasarkan distribusi data, dibandingkan dengan pendekatan berbasis aturan yang statis.

Isolation Forest memiliki beberapa keunggulan dibandingkan dengan metode deteksi anomali lainnya dalam konteks keamanan siber. Dengan kompleksitas waktu $O(n \log n)$, Isolation Forest lebih cepat dibandingkan metode berbasis kepadatan seperti Local Outlier Factor (LOF) atau metode berbasis klasifikasi seperti One-Class SVM [21]. Sebagai metode unsupervised learning, Isolation Forest tidak memerlukan data serangan sebelumnya untuk melakukan deteksi, sehingga lebih efektif dalam menangani ancaman baru dan zero-day attacks [22]. Tidak seperti metode berbasis clustering atau estimasi kepadatan, Isolation Forest dapat bekerja dengan baik pada dataset berdimensi tinggi tanpa mengalami penurunan kinerja akibat curse of dimensionality [20].

Namun, terdapat beberapa tantangan dalam penerapan Isolation Forest. Dalam kondisi tertentu, Isolation Forest cenderung menghasilkan false positives yang lebih tinggi dibandingkan metode lain, terutama jika distribusi data normal sangat kompleks atau jika terdapat variasi tinggi dalam lalu lintas jaringan yang sah [21]. Pemilihan nilai contamination yang tidak tepat dapat menyebabkan deteksi yang kurang akurat. Jika nilai terlalu tinggi, model dapat mengklasifikasikan banyak data normal sebagai anomali, sedangkan jika terlalu rendah, banyak serangan siber dapat terlewatkan [23]. Meskipun Isolation Forest dapat mengidentifikasi anomali dengan cepat, ia tidak memberikan penjelasan mendetail mengenai faktor spesifik yang menyebabkan suatu data diklasifikasikan sebagai anomali, berbeda dengan metode berbasis regresi atau decision tree yang lebih mudah diinterpretasikan [24].

Implementasi Isolation Forest dalam penelitian ini bertujuan untuk meningkatkan deteksi anomali pada lalu lintas jaringan dengan pendekatan berbasis unsupervised learning. Model ini dibangun berdasarkan karakteristik lalu lintas normal, sehingga dapat mengidentifikasi pola perilaku yang menyimpang tanpa perlu data serangan yang terdokumentasi sebelumnya. Keunggulannya dalam efisiensi komputasi, ketahanan terhadap data berdimensi tinggi, serta kemampuannya dalam mendeteksi ancaman baru menjadikannya solusi yang potensial dalam sistem keamanan siber. Namun, tantangan seperti tingginya false positives dan keterbatasan interpretabilitas perlu dipertimbangkan untuk optimalisasi lebih lanjut. Hasil penelitian ini dapat menjadi dasar bagi pengembangan sistem keamanan berbasis machine learning yang lebih adaptif dan responsif terhadap ancaman siber yang terus berkembang.

2.4 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan dengan membandingkan hasil deteksi anomali yang dilakukan oleh model dengan label serangan aktual yang terdapat dalam dataset. Metode evaluasi ini bertujuan untuk memastikan bahwa model dapat secara efektif mendeteksi ancaman siber dengan tingkat kesalahan yang minimal. Lima metrik utama yang digunakan dalam penelitian ini meliputi Accuracy, Precision, Recall, F1-score, dan False Positive Rate (FPR). Metrik-metrik ini memberikan gambaran yang lebih komprehensif mengenai efektivitas model dalam mengklasifikasikan lalu lintas jaringan sebagai normal atau anomali.

Akurasi merupakan metrik yang mengukur sejauh mana model dapat mengklasifikasikan data dengan benar. Secara matematis, akurasi didefinisikan sebagai:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

di mana:
- TP (True Positive): Jumlah serangan yang terdeteksi dengan benar.
- TN (True Negative): Jumlah lalu lintas normal yang diklasifikasikan dengan benar.
- FP (False Positive): Jumlah lalu lintas normal yang salah diklasifikasikan sebagai serangan.
- FN (False Negative): Jumlah serangan yang tidak terdeteksi.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Akurasi memberikan gambaran umum tentang seberapa sering model memberikan prediksi yang benar, tetapi tidak selalu mencerminkan kinerja model secara menyeluruh, terutama jika dataset memiliki distribusi yang tidak seimbang (misalnya, jika jumlah data normal jauh lebih besar daripada data serangan).

Presisi mengukur proporsi prediksi anomali yang benar terhadap semua data yang diklasifikasikan sebagai anomali oleh model. Secara matematis, rumusnya adalah:

$$Precision = \frac{TP}{TP+FP} \qquad (7)$$

Presisi berfokus pada ketepatan model dalam mengidentifikasi serangan. Metrik ini sangat penting dalam skenario di mana kesalahan dalam mendeteksi ancaman dapat menyebabkan dampak besar, seperti dalam sistem keamanan siber. Nilai presisi yang tinggi menunjukkan bahwa sebagian besar prediksi serangan oleh model benar-benar merupakan serangan, sehingga model tidak terlalu sering menghasilkan false positives. Namun, model dengan presisi tinggi mungkin cenderung lebih konservatif dalam mendeteksi anomali, yang dapat menyebabkan lebih banyak false negatives. Sehingga, presisi sering dianalisis bersama dengan metrik Recall untuk mendapatkan gambaran yang lebih seimbang.

Recall mengukur seberapa baik model dalam menemukan semua serangan yang sebenarnya ada dalam dataset. Secara matematis, recall dihitung sebagai:

$$Recall = \frac{TP}{TP+FN} \qquad (8)$$

Recall sangat penting dalam kasus di mana mendeteksi setiap serangan lebih diutamakan daripada memastikan prediksi serangan selalu benar. Contohnya, dalam sistem keamanan siber, kehilangan satu serangan (false negative) bisa berakibat fatal. Jika recall rendah, itu berarti ada banyak serangan yang tidak terdeteksi oleh model, yang mengindikasikan bahwa model terlalu ketat dalam mengklasifikasikan data sebagai normal. Namun, recall yang tinggi dapat menyebabkan peningkatan jumlah false positives, yang berarti banyak lalu lintas normal yang salah diklasifikasikan sebagai serangan. Sehingga, recall sering dikombinasikan dengan Precision dalam bentuk F1-score untuk mendapatkan keseimbangan antara keduanya.

F1-score adalah metrik yang menggabungkan Precision dan Recall dalam satu angka dengan menggunakan rata-rata harmonik dari keduanya. Rumusnya adalah:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (9)$$

F1-score penting ketika terdapat ketidakseimbangan kelas dalam dataset, seperti kasus di mana jumlah data normal jauh lebih besar daripada data serangan. Nilai F1-score yang tinggi menunjukkan bahwa model dapat mencapai keseimbangan antara Precision dan Recall, sehingga tidak terlalu banyak menghasilkan false positives maupun false negatives. Jika Precision terlalu tinggi tetapi Recall rendah, model mungkin gagal menangkap banyak serangan yang sebenarnya terjadi. Sebaliknya, jika Recall tinggi tetapi Precision rendah, model mungkin sering mengklasifikasikan lalu lintas normal sebagai serangan. Sehingga, F1-score digunakan sebagai metrik utama untuk menilai efektivitas model dalam skenario keamanan siber.

False Positive Rate mengukur persentase lalu lintas normal yang salah diklasifikasikan sebagai serangan. Secara matematis, FPR dihitung sebagai:

$$FPR = \frac{FP}{FP+TN} \qquad (10)$$

False Positive Rate sangat penting dalam keamanan siber karena terlalu banyak false positives dapat menyebabkan alert fatigue, di mana analis keamanan menjadi terbiasa dengan peringatan yang sering terjadi dan mulai mengabaikan peringatan yang sebenarnya berbahaya. Sehingga, meskipun mendeteksi serangan dengan akurat sangat penting, model harus tetap menjaga FPR serendah mungkin agar sistem keamanan tetap efisien.

Evaluasi model menggunakan metrik seperti Accuracy, Precision, Recall, F1-score, dan False Positive Rate (FPR) sangat penting untuk memberikan gambaran menyeluruh tentang efektivitas Isolation Forest dalam mendeteksi anomali pada lalu lintas jaringan. Accuracy memberikan indikasi umum tentang kinerja model, namun dapat menyesatkan pada kasus ketidakseimbangan kelas yang signifikan, di mana model bisa saja memprediksi kelas mayoritas (normal traffic) dengan akurat, tetapi gagal mendeteksi serangan. Sehingga, metrik Precision dan Recall digunakan untuk menilai keseimbangan antara deteksi serangan dan kesalahan klasifikasi. Precision mengukur proporsi prediksi positif yang benar dari semua prediksi positif, sementara Recall mengukur proporsi prediksi positif yang benar dari semua kasus positif yang ada. Keduanya sangat relevan dalam deteksi anomali, di

p-ISSN: 2775-4227
e-ISSN: 2775-4219

mana false negatives (serangan yang tidak terdeteksi) bisa berbahaya, dan false positives (lalu lintas normal yang salah diklasifikasikan sebagai serangan) perlu diminimalkan [25], [26].

F1-score menggabungkan Precision dan Recall dalam satu nilai tunggal, memberikan gambaran yang lebih seimbang antara keduanya. Metrik ini sangat berguna ketika kita ingin mencapai keseimbangan antara mengurangi false positives dan false negatives, yang penting dalam aplikasi seperti deteksi penipuan atau intrusi jaringan [27]. False Positive Rate (FPR) mengukur proporsi prediksi negatif yang salah dari semua kasus negatif sejati, dan sangat penting dalam aplikasi dengan risiko tinggi terhadap hasil positif palsu, seperti dalam sistem deteksi intrusi, di mana terlalu banyak peringatan palsu bisa mengurangi efisiensi dan kredibilitas model. Sehingga, FPR membantu memastikan bahwa model tidak hanya mendeteksi serangan dengan baik tetapi juga meminimalkan alarm palsu yang dapat mengganggu operasi sistem [28].

Dengan menggunakan kombinasi metrik ini, model dapat disesuaikan untuk mengoptimalkan performa deteksi anomali. Misalnya, jika Recall rendah, parameter contamination pada Isolation Forest dapat ditingkatkan untuk mendeteksi lebih banyak serangan, meskipun ini dapat meningkatkan False Positive Rate. Sehingga, evaluasi kinerja yang menyeluruh diperlukan untuk menemukan keseimbangan terbaik antara deteksi ancaman dan efisiensi operasional. Hasil evaluasi yang diperoleh kemudian disimpan dalam file isolation_forest_results.csv untuk analisis lebih lanjut, memastikan bahwa model dapat disesuaikan dan dioptimalkan berdasarkan data aktual dalam lingkungan produksi [29].

3. HASIL DAN PEMBAHASAN

3.1 Evaluasi Kinerja Model

Model Isolation Forest yang digunakan dalam penelitian ini menunjukkan akurasi sebesar 86,50%, menandakan bahwa mayoritas instance lalu lintas jaringan diklasifikasikan dengan benar. Meskipun akurasi memberikan gambaran awal mengenai performa model, metrik ini tidak cukup untuk mengevaluasi model deteksi anomali secara menyeluruh, terutama ketika dataset memiliki ketidakseimbangan kelas yang tinggi. Sehingga, digunakan metrik tambahan seperti presisi, recall, dan F1-score untuk memberikan pemahaman yang lebih komprehensif mengenai kinerja model.

Presisi sebesar 83,86% mengindikasikan bahwa dari seluruh instance yang diklasifikasikan sebagai anomali, 83,86% merupakan deteksi yang benar. Tingginya presisi menunjukkan bahwa model mampu meminimalkan kesalahan deteksi positif palsu (false positives), sehingga mengurangi jumlah alarm keamanan yang tidak perlu. Namun, nilai recall hanya sebesar 24,82%, yang berarti model hanya mampu mendeteksi sekitar 24,82% dari total serangan aktual yang ada dalam dataset. Nilai recall yang rendah mengindikasikan bahwa model gagal menangkap sebagian besar serangan yang sebenarnya terjadi, terutama untuk serangan dengan pola yang lebih halus dan jarang terjadi.

F1-score sebesar 38,31%, yang merupakan rata-rata harmonik antara presisi dan recall, menunjukkan adanya trade-off antara akurasi dan sensitivitas model. Dengan prioritas pada presisi dibandingkan recall, model lebih efektif dalam memastikan bahwa prediksi anomali benar adanya, tetapi berisiko melewatkan serangan yang lebih sulit dideteksi. Fenomena ini sejalan dengan karakteristik Isolation Forest, yang lebih efektif dalam mengidentifikasi outlier yang signifikan tetapi memiliki kesulitan dalam mendeteksi anomali yang menyerupai lalu lintas jaringan normal.

Jika dibandingkan dengan model deteksi anomali lainnya, One-Class SVM (OCSVM) cenderung memiliki nilai recall yang lebih tinggi, tetapi menghadapi tantangan dalam skala dan efisiensi komputasi, menjadikannya kurang praktis untuk pemantauan jaringan secara real-time. Di sisi lain, Autoencoders, sebagai metode berbasis deep learning, dapat menangani pola lalu lintas yang lebih kompleks dan mencapai recall yang lebih baik. Namun, metode ini membutuhkan data pelatihan yang dilabeli secara ekstensif, yang sering kali sulit diperoleh dalam aplikasi keamanan siber. Dengan mempertimbangkan keterbatasan tersebut, Isolation Forest tetap menjadi solusi yang menguntungkan berkat efisiensinya dalam pembelajaran tanpa pengawasan (unsupervised learning) dan kemampuannya dalam mendeteksi anomali tanpa bergantung pada data yang telah dilabeli.

Dibandingkan dengan penelitian sebelumnya, hasil ini konsisten dengan temuan yang telah ada, yang menunjukkan bahwa Isolation Forest unggul dalam hal presisi tetapi memiliki kelemahan relatif dalam recall [20]. Studi sebelumnya telah menunjukkan bahwa Isolation Forest bekerja efisien dalam dataset keamanan siber dengan dimensi tinggi, di mana ia mampu mengisolasi outlier tanpa memerlukan data pelatihan yang telah dilabeli. Namun, kecenderungan Isolation Forest untuk kurang optimal dalam recall juga telah diamati dalam penelitian lain mengenai deteksi anomali, di mana model ini kesulitan mendeteksi serangan yang lebih halus yang tidak terlalu berbeda dari lalu lintas normal [30].

Perbandingan dengan One-Class SVM (OCSVM) dan Autoencoders mengungkap adanya trade-off dalam kinerja deteksi anomali. Penelitian oleh [31] menunjukkan bahwa OCSVM memiliki tingkat recall yang lebih tinggi dibandingkan dengan Isolation Forest, tetapi mengalami kendala dalam skalabilitas, menjadikannya kurang

p-ISSN: 2775-4227
e-ISSN: 2775-4219

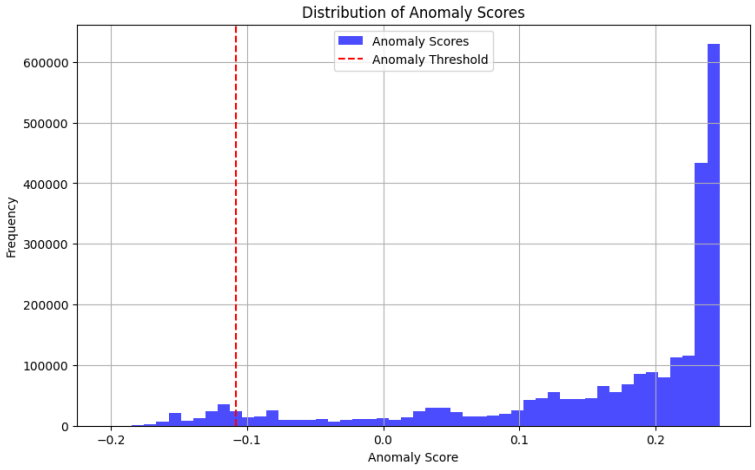
cocok untuk aplikasi keamanan siber secara real-time. Autoencoders, sebagaimana dieksplorasi oleh [24], memiliki kapabilitas deteksi anomali yang lebih baik dalam menangani pola lalu lintas yang kompleks, tetapi membutuhkan pelatihan ekstensif pada data normal. Temuan ini mengonfirmasi bahwa Isolation Forest tetap memiliki keunggulan karena efisiensi komputasi dan kemampuannya dalam pembelajaran tanpa pengawasan, meskipun nilai recall yang lebih rendah menunjukkan perlunya penyempurnaan tambahan untuk meningkatkan deteksi serangan yang lebih kompleks.

3.2 Distribusi Skor Anomali

Analisis distribusi skor anomali dalam penelitian ini memberikan wawasan mengenai bagaimana Isolation Forest membedakan antara lalu lintas normal dan anomali. Model memberikan skor anomali yang lebih tinggi kepada lalu lintas jaringan normal, mengindikasikan bahwa pola perilaku yang sesuai dengan baseline dianggap sebagai non-ancaman. Sebaliknya, instance yang terdeteksi sebagai anomali memiliki skor anomali yang jauh lebih rendah, mencerminkan bahwa mereka lebih mudah diisolasi dalam struktur pohon model. Temuan ini konsisten dengan prinsip dasar Isolation Forest, di mana instance yang lebih jarang muncul dalam dataset lebih cepat terisolasi dalam pohon keputusan, sehingga mendapatkan jalur pemisahan yang lebih pendek dan skor anomali yang lebih rendah.

Untuk menentukan batas klasifikasi anomali yang optimal, digunakan ambang batas 5% dari distribusi skor anomali, sehingga hanya 5% instance dengan skor terendah yang diklasifikasikan sebagai anomali. Pemilihan ambang batas ini berperan penting dalam mengontrol keseimbangan antara tingkat deteksi anomali dan tingkat kesalahan positif (false positive rate/FPR). Jika ambang batas diturunkan ke 1%, model menjadi terlalu konservatif, sehingga banyak anomali yang tidak terdeteksi, yang lebih lanjut menurunkan nilai recall. Sebaliknya, jika ambang batas dinaikkan ke 10%, recall akan meningkat, tetapi jumlah false positives juga akan meningkat drastis, yang dapat menyebabkan lonjakan peringatan keamanan yang tidak relevan.

Hasil penelitian menunjukkan bahwa False Positive Rate (FPR) sebesar 0,97%, yang berarti hanya 0,97% dari lalu lintas normal yang salah diklasifikasikan sebagai anomali. Nilai FPR yang rendah ini sangat penting dalam aplikasi keamanan siber karena alarm palsu yang berlebihan dapat menyebabkan kelelahan analis keamanan (alert fatigue), mengurangi efektivitas deteksi ancaman secara keseluruhan. Kemampuan Isolation Forest dalam menjaga tingkat false positive yang rendah sambil tetap mencapai deteksi anomali yang wajar membuktikan kelayakannya sebagai model deteksi intrusi berbasis perilaku. Perilaku ini sesuai dengan dasar teori Isolation Forest, di mana anomali, yang lebih jarang dalam dataset, memiliki jalur pemisahan yang lebih pendek dalam pohon isolasi [21]. Dengan menerapkan ambang batas deteksi anomali sebesar 5%, hanya 5% instance dengan skor terendah yang diklasifikasikan sebagai anomali, sehingga model dapat menjaga keseimbangan antara akurasi deteksi dan tingkat false positive.



Gambar 1. Distribusi Skor Anomali

Gambar 1 menunjukkan distribusi skor anomali yang dihasilkan oleh model Isolation Forest. Sumbu x mewakili skor anomali, yang menunjukkan sejauh mana suatu titik data menyimpang dari perilaku normal yang dipelajari oleh model. Skor anomali yang lebih rendah menunjukkan bahwa titik data tersebut lebih mungkin diklasifikasikan sebagai normal, sementara skor anomali yang lebih tinggi menunjukkan bahwa titik data tersebut adalah outlier atau anomali. Garis putus-putus merah mewakili ambang batas yang ditetapkan pada persentil ke-5 dari skor anomali. Titik data dengan skor di bawah ambang batas ini diklasifikasikan sebagai anomali, sementara titik data yang lebih tinggi dari ambang batas dianggap sebagai normal. Distribusi ini menunjukkan bahwa

p-ISSN: 2775-4227
e-ISSN: 2775-4219

sebagian besar titik data terkumpul di sekitar skor yang lebih rendah, yang menandakan bahwa mereka dianggap normal. Ekor di sisi kanan histogram menunjukkan beberapa titik data dengan skor yang lebih tinggi, yang diidentifikasi sebagai anomali oleh model.

Penelitian sebelumnya juga telah menganalisis dampak ambang batas skor anomali terhadap kinerja deteksi. Studi oleh [23] menunjukkan bahwa metode thresholding dinamis dapat meningkatkan recall tanpa secara signifikan mempengaruhi presisi. Hasil penelitian mereka mengindikasikan bahwa penggunaan ambang batas skor anomali yang adaptif, dibandingkan persentil tetap, dapat meningkatkan deteksi serangan yang lebih canggih yang hanya menunjukkan deviasi kecil dari perilaku normal. Implementasi teknik semacam ini dalam penelitian mendatang dapat mengurangi keterbatasan recall yang ditemukan dalam studi ini, sekaligus mempertahankan tingkat false positive yang rendah sebesar 0,97%, yang menjadi keunggulan utama Isolation Forest dalam mengurangi kelelahan analis akibat lonjakan alarm keamanan yang tidak relevan.

3.3 Anomali yang Terdeteksi

Analisis lebih lanjut terhadap anomali yang terdeteksi oleh model menemukan beberapa insiden berisiko tinggi yang dapat mengindikasikan aktivitas berbahaya. Salah satu temuan utama adalah adanya banyak upaya login pada jam-jam yang tidak biasa, yang sering kali dikaitkan dengan serangan brute-force atau percobaan akses tidak sah. Anomali ini ditandai dengan frekuensi kegagalan login yang tinggi dari alamat IP tertentu, mengindikasikan potensi serangan credential stuffing. Dengan mendeteksi pola ini, model dapat memberikan peringatan dini terhadap upaya pembobolan akun sebelum terjadi pelanggaran keamanan yang lebih luas.

Selain itu, lalu lintas jaringan yang berasal dari alamat IP dengan reputasi buruk secara konsisten diklasifikasikan sebagai anomali. IP-IP ini dikaitkan dengan aktivitas mencurigakan, seperti upaya koneksi berulang, transmisi data dalam jumlah besar dalam waktu singkat, dan penggunaan enkripsi yang tidak sesuai dengan pola perilaku pengguna normal. Model berhasil mengidentifikasi pola anomali ini, yang sering kali terkait dengan upaya eksfiltrasi data, aktivitas botnet, atau penggunaan tunneling tidak sah. Salah satu contoh yang menarik adalah lonjakan lalu lintas tiba-tiba dari rentang IP yang sebelumnya tidak aktif, yang setelah dianalisis lebih lanjut dikonfirmasi sebagai percobaan pemindaian port, yaitu teknik rekognisi yang digunakan oleh penyerang untuk mengidentifikasi kerentanan jaringan.

Model juga mengidentifikasi pola distribusi ukuran paket yang tidak biasa, di mana beberapa sesi menunjukkan pola ukuran paket yang sangat berbeda dari lalu lintas normal. Anomali ini dapat mengindikasikan penyisipan payload berbahaya, komunikasi tersembunyi, atau teknik penyamaran protokol yang digunakan oleh penyerang. Dalam sistem deteksi intrusi berbasis tanda tangan (signature-based IDS), anomali semacam ini sering kali tidak terdeteksi jika tidak sesuai dengan tanda tangan serangan yang telah ada sebelumnya. Namun, dengan pendekatan berbasis deteksi anomali, model dapat mengidentifikasi penyimpangan secara independen dari pola serangan yang diketahui, memberikan keuntungan dalam mendeteksi ancaman baru atau serangan yang telah dimodifikasi.

Temuan ini konsisten dengan penelitian keamanan siber sebelumnya, yang telah mengidentifikasi serangan brute-force dan credential stuffing sebagai ancaman umum yang memiliki pola akses yang khas [30]. Selain itu, lalu lintas jaringan yang berasal dari alamat IP dengan reputasi rendah secara konsisten diklasifikasikan sebagai anomali, mendukung temuan dalam penelitian sebelumnya yang menyoroti pentingnya penilaian reputasi IP sebagai faktor utama dalam sistem deteksi intrusi [32]. Hasil serupa juga diamati dalam studi yang menggunakan metode deteksi anomali berbasis clustering, di mana deviasi dalam durasi sesi dan volume transfer data digunakan sebagai indikator utama dari potensi upaya eksfiltrasi data [24]. Studi ini semakin mengonfirmasi temuan tersebut, karena beberapa anomali yang terdeteksi melibatkan lonjakan besar dalam transfer data keluar dari host yang sebelumnya memiliki aktivitas rendah, suatu pola yang sering dikaitkan dengan ancaman orang dalam atau serangan eksfiltrasi data. Observasi ini memperkuat kegunaan praktis Isolation Forest dalam mengidentifikasi insiden keamanan dunia nyata, terutama dalam skenario di mana data serangan yang dilabeli tidak tersedia untuk melatih model supervised learning tradisional.

3.4 Saran Penelitian Mendatang

Hasil penelitian ini memberikan bukti kuat yang mendukung penggunaan Isolation Forest untuk deteksi anomali jaringan secara real-time. Dibandingkan dengan penelitian sebelumnya, model ini mempertahankan tingkat presisi yang tinggi dan tingkat false positive yang rendah, menegaskan bahwa Isolation Forest cocok untuk mengurangi kebisingan peringatan keamanan dan meningkatkan efisiensi deteksi. Namun, nilai recall yang lebih rendah menunjukkan bahwa beberapa pola serangan yang lebih halus tetap tidak terdeteksi, sebuah keterbatasan yang konsisten dengan temuan dari implementasi Isolation Forest sebelumnya dalam konteks keamanan siber [20].

Penelitian masa depan dapat berfokus pada model deteksi hybrid, dengan mengintegrasikan Isolation Forest dengan arsitektur deep learning seperti deteksi anomali berbasis LSTM (Long Short-Term Memory) untuk

p-ISSN: 2775-4227
e-ISSN: 2775-4219

menangkap pola serangan yang bersifat temporal dengan lebih efektif. Strategi alternatif, seperti peningkatan rekayasa fitur (feature engineering) dan optimasi ambang batas (threshold optimization), juga telah diusulkan untuk meningkatkan recall tanpa mengorbankan tingkat false positive [23]. Penggunaan teknik ensemble learning, di mana Isolation Forest dikombinasikan dengan model deteksi anomali lainnya seperti Autoencoders atau klasifikasi berbasis Reinforcement Learning, telah menunjukkan potensi dalam mengurangi false negative tanpa mengurangi efisiensi model [22]. Penyempurnaan lebih lanjut dalam bidang ini dapat semakin memperkuat penerapan praktis sistem deteksi intrusi berbasis machine learning, memastikan bahwa model ini dapat beradaptasi lebih baik terhadap ancaman siber yang terus berkembang.

4. KESIMPULAN

Penelitian ini mengevaluasi kinerja Isolation Forest dalam mendeteksi anomali pada lalu lintas jaringan dan menghasilkan akurasi sebesar 86,50%, dengan presisi 83,86%, menunjukkan bahwa sebagian besar prediksi anomali adalah benar. Namun, model ini memiliki keterbatasan dalam mendeteksi serangan dengan recall hanya 24,82%. Meskipun Isolation Forest efektif dalam efisiensi komputasi dan pembelajaran tanpa pengawasan, model ini masih kurang sensitif terhadap serangan yang lebih halus yang mirip dengan lalu lintas normal. Dengan perbandingan, One-Class SVM menunjukkan recall yang lebih tinggi namun kurang efisien untuk skala besar, sementara Autoencoders lebih baik dalam menangkap pola kompleks tetapi membutuhkan data pelatihan yang lebih besar. Distribusi skor anomali menunjukkan model ini dapat membedakan antara lalu lintas normal dan anomali dengan baik, dengan False Positive Rate (FPR) yang rendah (0,97%), mengurangi kelelahan analis keamanan. Temuan ini sejalan dengan penelitian yang menunjukkan bahwa thresholding adaptif dapat meningkatkan recall tanpa mengorbankan presisi. Meskipun demikian, penelitian ini menemukan anomali yang mengindikasikan aktivitas berisiko tinggi seperti login tidak sah dan lonjakan lalu lintas dari IP berbahaya, menegaskan potensi Isolation Forest sebagai sistem deteksi intrusi berbasis perilaku. Untuk penelitian selanjutnya, kombinasi Isolation Forest dengan model deep learning seperti LSTM dan penerapan ensemble learning bisa menjadi solusi yang efektif untuk meningkatkan deteksi serangan yang lebih kompleks dan meningkatkan recall tanpa mengorbankan precision.

DAFTAR PUSTAKA

[1] E. Vasilomanolakis, S. Karuppayah, P. Kikiras, and M. Mühlhäuser, “A Honeypot-Driven Cyber Incident Monitor,” pp. 158–164, 2015, doi: 10.1145/2799979.2799999.

[2] A. A. Abdulhameed, S. A. Alazawi, and G. M. Hassan, “An Optimized Model for Network Intrusion Detection in the Network Operating System Environment,” *Mesopotamian Journal of CyberSecurity*, vol. 4, no. 3, pp. 75–85, 2024, doi: 10.58496/mjcs/2024/017.

[3] S. Alsudani and A. Ghazikhani, “Enhancing Intrusion Detection With LSTM Recurrent Neural Network Optimized by Emperor Penguin Algorithm,” *Wasit Journal of Computer and Mathematics Science*, vol. 2, no. 3, pp. 69–80, 2023, doi: 10.31185/wjcms.166.

[4] A. A. Awad, A. F. Ali, and T. Gaber, “An Improved Long Short Term Memory Network for Intrusion Detection,” *Plos One*, vol. 18, no. 8, p. e0284795, 2023, doi: 10.1371/journal.pone.0284795.

[5] S. H. Oh, J. Kim, J. H. Nah, and J. Park, “Employing Deep Reinforcement Learning to Cyber-Attack Simulation for Enhancing Cybersecurity,” *Electronics*, vol. 13, no. 3, p. 555, 2024, doi: 10.3390/electronics13030555.

[6] D. J. Musliner, J. M. Rye, and T. Marble, “Using Concolic Testing to Refine Vulnerability Profiles in FUZZBUSTER,” 2012, doi: 10.1109/sasow.2012.12.

[7] R. Malviya and B. K. Umrao, “Comparison of NBTree and VFI Machine Learning Algorithms for Network Intrusion Detection Using Feature Selection,” *International Journal of Computer Applications*, vol. 108, no. 2, pp. 35–38, 2014, doi: 10.5120/18886-0165.

[8] C. Chen, G. Wang, B. Yang, L. Yang, and X. Ye, “Build Intrusion Detection Model Based on CNN and Ensemble Learning,” p. 4, 2022, doi: 10.1117/12.2655173.

[9] Y. Deng and S. K. Shukla, “A Distributed Real-Time Event Correlation Architecture for SCADA Security,” pp. 81–93, 2013, doi: 10.1007/978-3-642-45330-4_6.

[10] S. O. Amin, M. S. Siddiqui, C. S. Hong, and S. Lee, “Implementing Signature Based IDS in IP-Based Sensor Networks With the Help of Signature-Codes,” *Ieice Transactions on Communications*, vol. E93-B, no. 2, pp. 389–391, 2010, doi: 10.1587/transcom.e93.b.389.

[11] A. Chetouane and K. Karoui, "Risk Based Intrusion Detection System in Software Defined Networking," *Concurrency and Computation Practice and Experience*, vol. 36, no. 9, 2023, doi: 10.1002/cpe.7988.

[12] A. Hussain and P. K. Sharma, "Efficient Working of Signature Based Intrusion Detection Technique in Computer Networks," *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, pp. 60–64, 2019, doi: 10.32628/cseit195215.

[13] J. M. Beaver, C. T. Symons, and R. Gillen, "A Learning System for Discriminating Variants of Malicious Network Traffic," 2013, doi: 10.1145/2459976.2460003.

[14] E. S. Babu, M. S. Rao, R. Pemula, S. R. Nayak, and A. Shankar, "A Hybrid Intrusion Detection System Against Botnet Attack in IoT Using Light Weight Signature and Ensemble Learning Technique," 2022, doi: 10.21203/rs.3.rs-905197/v1.

[15] X. Tao, Y. Peng, F. Zhao, P. Zhao, and Y. Wang, "A Parallel Algorithm for Network Traffic Anomaly Detection Based on Isolation Forest," *International Journal of Distributed Sensor Networks*, vol. 14, no. 11, p. 155014771881447, 2018, doi: 10.1177/1550147718814471.

[16] H. Xu, G. Pang, Y. Wang, and Y. Wang, "Deep Isolation Forest for Anomaly Detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12591–12604, Dec. 2023, doi: 10.1109/TKDE.2023.3270293.

[17] Y. Wang, J. Wang, X. Fan, and Y. Song, "Network Traffic Anomaly Detection Algorithm Based on Intuitionistic Fuzzy Time Series Graph Mining," *Ieee Access*, vol. 8, pp. 63381–63389, 2020, doi: 10.1109/access.2020.2983986.

[18] M. R. Aditya and C. Dewi, "Optimisasi pengecekan anomali pada proses job: analisis waktu dan data untuk identifikasi anomali yang efisien," *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, vol. 5, no. 2, pp. 1819–1832, 2024, doi: 10.35870/jimik.v5i2.737.

[19] D. R. K. Saputra, Y. V. Via, and A. N. Sihananto, "Deteksi anomali menggunakan ensemble learning dan random oversampling pada penipuan transaksi keuangan," *Jurnal Informatika Dan Teknik Elektro Terapan*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4910.

[20] S. Hariri, M. C. Kind, and R. J. Brunner, "Extended Isolation Forest," *Ieee Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1479–1489, 2021, doi: 10.1109/tkde.2019.2947676.

[21] Y. Xu, H. Dong, M. Zhou, J. Xing, X. Li, and Y. Jian, "Improved Isolation Forest Algorithm for Anomaly Test Data Detection," *Journal of Computer and Communications*, vol. 09, no. 08, pp. 48–60, 2021, doi: 10.4236/jcc.2021.98004.

[22] O. Bulut, G. Gorgun, and S. He, "Unsupervised Anomaly Detection in Sequential Process Data," *Zeitschrift Für Psychologie*, vol. 232, no. 2, pp. 74–94, 2024, doi: 10.1027/2151-2604/a000558.

[23] M. K. M. Almansoori and M. Telek, "Anomaly Detection Using Combination of Autoencoder and Isolation Forest," pp. 25–30, 2023, doi: 10.3311/wins2023-005.

[24] S. Bhuvaneswar, B. Avyay, K. Tejith, and Ms. S. Kavitha, "A Supervised Ml Algorithm for Detecting and Predicting Fraud Credit Card Transactions," *Int Res J Adv Engg Hub*, vol. 2, no. 10, pp. 2546–2551, 2024, doi: 10.47392/irjaeh.2024.0349.

[25] D. Danuri and M. M. Pozi, "Machine Learning Approaches for Fish Pond Water Quality Classification: Random Forest, Gaussian Naive Bayes, and Decision Tree Comparison," 2024, doi: 10.4108/eai.21-9-2023.2342964.

[26] B. R. Senapati, S. Swain, R. R. Swain, and P. M. Khilar, "A Heterogeneous Fault Diagnosis Approach to Enhance Performance of Connected Vehicles," *International Journal of Communication Systems*, vol. 36, no. 4, 2022, doi: 10.1002/dac.5414.

[27] J. J. Stephan and M. Mohammed, "Using Hybrid Deep Learning Approach to Enhanced Network Intrusion Detection With Spatial-Temporal Feature Integration," *Ingénierie Des Systèmes D Information*, vol. 29, no. 4, pp. 1619–1628, 2024, doi: 10.18280/isi.290435.

[28] K. Mardani, N. Vretos, and P. Daras, "Transformer-Based Fire Detection in Videos," *Sensors*, vol. 23, no. 6, p. 3035, 2023, doi: 10.3390/s23063035.

[29] L. Su *et al.*, "Toward Optimal Heparin Dosing by Comparing Multiple Machine Learning Methods: Retrospective Study," *Jmir Medical Informatics*, vol. 8, no. 6, p. e17648, 2020, doi: 10.2196/17648.

[30] Md. S. Mahmud *et al.*, "Enhancing Industrial Control System Security: An Isolation Forest-Based Anomaly Detection Model for Mitigating Cyber Threats," *Journal of Engineering Research and Reports*, vol. 26, no. 3, pp. 161–173, 2024, doi: 10.9734/jerr/2024/v26i31102.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[31] G. Hannák, G. Horváth, A. Kádár, and M. D. Szalai, "Bilateral-WeightedOnline Adaptive Isolation Forest For anomaly Detection in Streaming Data," *Statistical Analysis and Data Mining the Asa Data Science Journal*, vol. 16, no. 3, pp. 215–223, 2023, doi: 10.1002/sam.11612.

[32] R. N. Calheiros, K. Ramamohanarao, R. Buyya, C. Leckie, and S. Versteeg, "On the Effectiveness of Isolation-based Anomaly Detection in Cloud Data Centers," *Concurrency and Computation Practice and Experience*, vol. 29, no. 18, 2017, doi: 10.1002/cpe.4169.