COSMIC JURNAL TEKNIK
VOL. 2, NO. 3, AGUSTUS 2025: 88-95
DOI : 10.55537/cosmic
e-ISSN 3063-4687

Deteksi Trafik Anomali Berdasarkan Pola Trafik Menggunakan *Isolation Forest*

Muhammad 'Azam Al-Akbar[1], Ardan Pratama Y[2], Agil Naufal Al Habib Gurning[3]
[1,2,3]Fakultas Informatika, Telkom University
E-mail: [1]mazamakbar@student.telkomuniversity.ac.id,
[2]ardanpratamapro@student.telkomuniversity.ac.id, [3] agilnaufal@student.telkomuniversity.ac.id
Korespondensi : hesmiariayanti@telkomuniversity.ac.id

Abstrak

Peningkatan kompleksitas trafik jaringan di era digital menimbulkan tantangan dalam mendeteksi aktivitas anomali yang berpotensi membahayakan sistem. Penelitian ini mengusulkan pemanfaatan algoritme *Isolation Forest* sebagai metode deteksi anomali berbasis *unsupervised learning* untuk mengidentifikasi pola trafik yang menyimpang dari perilaku normal. Dataset yang digunakan adalah LUFlow, yaitu kumpulan data flow-level yang merepresentasikan trafik jaringan nyata yang telah dilabeli sebagai benign, malicious, dan outlier. Tahapan penelitian meliputi *preprocessing data*, standarisasi fitur, pelatihan model, visualisasi hasil, dan evaluasi performa menggunakan metrik *confusion matrix*, *precision*, *recall*, dan *F1-score*. Hasil eksperimen menunjukkan bahwa model berhasil mengidentifikasi trafik menyimpang dengan akurasi deteksi terhadap outlier sebesar 49%, namun belum efektif dalam mendeteksi serangan bot secara eksplisit. Visualisasi *scatter plot* memperkuat bahwa anomali terdistribusi jauh dari klaster trafik normal. Penelitian ini menegaskan potensi *Isolation Forest* dalam deteksi trafik anomali berbasis statistik, dan membuka peluang integrasi metode lanjutan seperti *autoencoder* atau *graph learning* untuk meningkatkan sensitivitas deteksi.

Kata kunci: Deteksi Anomali, Trafik Jaringan, Isolation Forest, LUFlow

Abstract

The increasing complexity of network traffic in the digital era presents challenges in detecting anomalous activities that may threaten system integrity. This study proposes the use of the Isolation Forest algorithm as an unsupervised learning-based anomaly detection method to identify traffic patterns that deviate from normal behavior. The research utilizes the LUFlow dataset, a flow-level network traffic dataset labeled as benign, malicious, and outlier. The methodological steps include data preprocessing, feature standardization, model training, result visualization, and performance evaluation using confusion matrix, precision, recall, and F1-score metrics. Experimental results indicate that the model successfully detected outliers with an accuracy of 49%, yet was not effective in explicitly identifying bot attacks. The scatter plot visualization confirms that anomalies are distributed outside the main cluster of normal traffic. This study highlights the potential of Isolation Forest in statistical-based traffic anomaly detection and suggests future integration with advanced techniques such as autoencoders or graph-based learning to improve detection sensitivity.

Keywords*: Anomaly Detection, Network Traffic, Isolation Forest, LUFlow*

1. PENDAHULUAN

Kemajuan teknologi informasi telah memungkinkan pertukaran data dalam jaringan komputer terjadi secara masif dan cepat. Namun, peningkatan konektivitas ini juga menimbulkan tantangan besar dalam menjaga keamanan jaringan. Salah satu ancaman yang signifikan adalah munculnya trafik anomali, yaitu lalu lintas jaringan yang menyimpang dari pola normal dan dapat mengindikasikan keberadaan aktivitas berbahaya seperti scanning, botnet, DDoS, maupun infeksi malware [1], [2], [3].

Trafik anomali umumnya tidak memiliki pola eksplisit, sehingga pendekatan berbasis rule (signature-based) menjadi kurang efektif dalam mendeteksinya [4]. Trafik anomali juga dapat dikenali melalui analisis statistik terhadap distribusi byte dalam paket atau flow, sebagaimana

© 2025 The Author(s). Published by AIRA . This is an open access article under the CC BY-SA license (http://creativecommons.org/licenses/by-sa/4.0/)

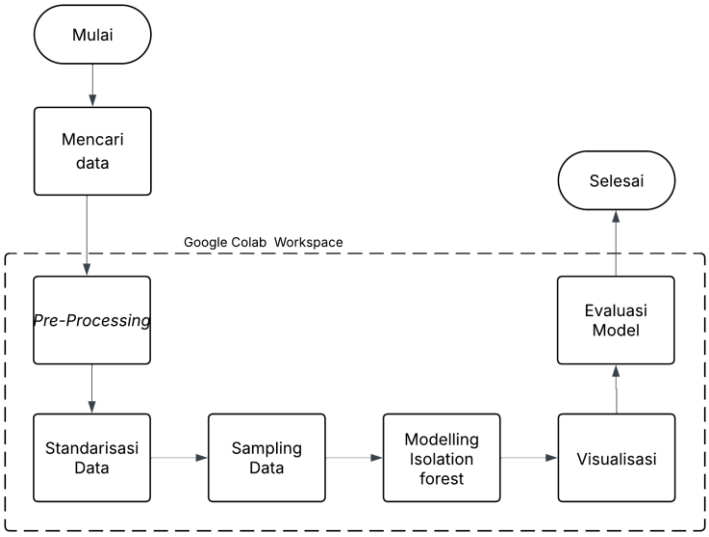
e-ISSN 3063-4687

dijelaskan dalam penelitian oleh Mahoney [5]. Oleh karena itu, metode deteksi berbasis anomali (anomaly-based detection) menjadi pendekatan yang lebih adaptif untuk mengenali pola baru yang belum diketahui sebelumnya. Salah satu teknik populer dalam kategori ini adalah algoritme *Isolation Forest* (iForest), yang bekerja dengan mengisolasi titik data berdasarkan kedalaman pohon acak, dan memisahkan data yang "mudah terisolasi" sebagai outlier [6], [7], [8].

Isolation Forest telah diterapkan dalam berbagai domain seperti industri manufaktur [3], optimisasi sistem campuran beton [7], dan deteksi anomali trafik jaringan berbasis entropi [9], [10]. Peningkatan performa juga dilaporkan ketika iForest digabungkan dengan metode *Local Outlier Factor* (LOF), sliding window, maupun deep ensemble [11], [12], [13]. Tantangan dalam penerapannya mencakup evaluasi performa pada data streaming, penyesuaian terhadap *concept drift*, dan kebutuhan akan seleksi fitur yang relevan [14], [15], [16].

Dataset LUFlow dari Lancaster University menjadi salah satu sumber data terbuka yang relevan untuk mendukung eksperimen dalam bidang ini. Dataset ini disusun menggunakan honeypot dan framework Citrus, dan mencerminkan trafik jaringan nyata yang dikategorikan sebagai benign, malicious, dan outlier [19].

2. METODE PENELITIAN



Gambar 1. *Flowchart* metode penelitian

Proses penelitian ini digambarkan dalam bentuk *flowchart* sebagaimana ditunjukkan pada Gambar 1. Flowchart ini merepresentasikan alur utama deteksi trafik anomali menggunakan algoritme *Isolation Forest*, yang terdiri dari beberapa tahapan: pencarian data, pra-pemrosesan, pemodelan, visualisasi hasil, dan evaluasi performa model.

Proses implementasi dilakukan menggunakan Google Colaboratory (Google Colab), yaitu sebuah platform pemrograman berbasis cloud yang mendukung eksekusi skrip Python dan kompatibel dengan pustaka seperti Pandas, Scikit-learn, Matplotlib, dan Seaborn. Platform ini dipilih karena menyediakan lingkungan pemrosesan yang fleksibel, efisien, serta terintegrasi dengan Google Drive untuk keperluan pengelolaan dataset.

Tahapan dimulai dengan pemuatan (loading) dataset lalu lintas jaringan dari Google Drive ke dalam Google Colab. Selanjutnya dilakukan pra-pemrosesan, yang mencakup konversi atribut kategorikal menjadi numerik, standarisasi fitur numerik, dan pengambilan sampel data sebanyak 10.000 baris secara acak untuk efisiensi proses. Data yang telah siap kemudian digunakan untuk melatih model *Isolation Forest* guna mengidentifikasi trafik anomali.

© 2025 The Author(s). Published by AIRA . This is an open access article under the CC BY-SA license (http://creativecommons.org/licenses/by-sa/4.0/)

e-ISSN 3063-4687

Hasil prediksi model divisualisasikan menggunakan grafik *scatter plot* untuk menunjukkan distribusi trafik normal dan anomali berdasarkan fitur utama. Evaluasi terhadap kinerja model dilakukan dengan membandingkan hasil deteksi terhadap label ground truth menggunakan metrik *confusion matrix*, *precision*, *recall*, dan *F1-score*. Proses ini dirancang untuk mempermudah eksperimen berbasis Python dalam platform berbasis cloud yang efisien dan fleksibel [4], [6], [14], [19].

2.1 Dataset

Penelitian ini menggunakan dataset LUFlow yang bersumber dari Kaggle [19]. LUFlow merupakan dataset berbasis flow-level yang dirancang untuk sistem deteksi intrusi jaringan. Dataset ini dikumpulkan melalui komposisi honeypot pada ruang alamat IP milik Lancaster University, dengan mekanisme pelabelan otomatis yang didukung oleh layanan threat intelligence.Label pada dataset terdiri dari trafik *benign* (normal), *malicious* (serangan), dan *outlier*, yaitu flow yang tidak bisa dikategorikan sebagai malicious maupun benign, namun menyimpang dari profil trafik normal. Label outlier ini disertakan untuk mendorong analisis lebih lanjut terhadap kemungkinan ancaman tersembunyi.

Dataset diperbarui secara kontinu menggunakan framework Citrus, dan disusun berdasarkan waktu pengambilan data. Data dalam penelitian ini berasal dari salah satu batch dataset LUFlow yang diambil pada bulan Juni 2022 (2022/06), dan berisi 590.087 entri trafik yang dikumpulkan menggunakan tool Joy milik Cisco. Adapun fitur seperti *dest_ip, dest_port, num_pkts_out, num_pkts_in, src_port, time_end, time_start, total_entry, label, duration* yang bisa dilihat pada gambar 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	avg_ipt	bytes_in	bytes_out	dest_ip	dest_port	entropy	num_pkts_out	num_pkts_in	proto	src_ip	src_port	time_end	time_start	total_entr	label	duration
2	500	14280	630	786	44010	2.289.381	30	60	6	786	9300	1,66E+15	1,66E+15	34134.67	benign	########
3	08.05	329	5559	786	9200	41.582	2	2	6	786	52136	1,66E+15	1,66E+15	24483.48	benign	0.017639
4	30.35.00	24575	27280	786	9300	1.766.307	129	117	6	786	52928	1,66E+15	1,66E+15	91591.84	benign	3.241.121
5	59	89	183	786	445	3.611.253	6	6	6	786	19237	1,66E+15	1,66E+15	9.822.609	outlier	0.24653
6	10.685.714.285.714.200	34	29	786	5900	5.139.898	11	10	6	786	47008	1,66E+15	1,66E+15	3.238.136	outlier	0.988713
7	9.957.142.857.142.850	34	29	786	5900	509.617	11	10	6	786	43036	1,66E+15	1,66E+15	########	outlier	0.972707
8	10.428.571.428.571.400	34	29	786	5900	5.127.916	10	10	6	786	39668	1,66E+15	1,66E+15	3.230.587	outlier	0.983964
9	69.05.00	19	43	786	3389	3.678.133	4	3	6	786	6765	1,66E+15	1,66E+14	########	outlier	0.241149
10	102	34	29	786	5900	5.020.696	11	10	6	786	45206	1,66E+15	1,66E+15	########	outlier	0.957799
11	0	8	8	786		03.00	1	1	1	786		1,66E+15	1,66E+15	########	outlier	5.9E-5
12	53.666.666.666.666.600	89	183	786	445	3.611.253	6	6	6	786	18655	1,66E+15	1,66E+15	9.822.609	outlier	0.237491
13	49.666.666.666.666.600	89	183	786	445	3.611.253	6	6	6	786	19454	1,66E+15	1,66E+15	9.822.609	outlier	0.220713
14	53.666.666.666.666.600	89	183	786	445	3.611.253	6	6	6	786	19286	1,66E+15	1,66E+15	9.822.609	outlier	0.231811
15	102	34	29	786	5900	5.056.022	11	10	6	786	39480	1,66E+15	1,66E+15	3.185.294	outlier	0.98447
16	0	8	8	786		03.00	1	1	1	786		1,66E+14	1,66E+15	########	outlier	5.9E-5
17	0	8	8	786		03.00	1	1	1	786		1,66E+15	1,66E+15	########	outlier	5.2E-5
18	0	8	8	786		0,135417	1	1	1	786		1,66E+15	1,66E+15	########	outlier	5.2E-5
19	0	8	8	786		03.00	1	1	1	786		1,66E+15	1,66E+13	########	outlier	5.4E-5
20	66	89	183	786	445	3.611.253	6	6	6	786	19234	1,66E+15	1,66E+15	9.822.609	outlier	0.300301
21	8.533.333.333.333.330	89	183	786	445	3.611.253	6	6	6	786	18965	1,66E+15	1,66E+14	9.822.609	outlier	0.375158
22	0	8	8	786		03.00	1	1	1	786		1,66E+15	1,66E+15	########	outlier	5.7E-5

2022.06.14

Gambar 2. Dataset Awal LUFlow

2.2 Prepocessing

Data yang digunakan dalam penelitian ini terdiri dari lebih dari 590.000 baris flow jaringan. Untuk efisiensi komputasi, dilakukan sampling sebanyak 10.000 baris secara acak menggunakan fungsi *sample()* dari pustaka pandas. Seluruh fitur bertipe kategori diubah ke dalam bentuk numerik, seperti protokol, label trafik, dan lainnya, dienkode ke dalam bentuk numerik menggunakan algoritme LabelEncoder dari scikit-learn. Proses ini bertujuan agar data dapat diproses oleh algoritme machine learning berbasis numerik [4], [6], [17].

2.3 Standarisasi Fitur

Sebelum diterapkan ke dalam model deteksi anomali, seluruh data yang telah di-encode kemudian distandarisasi menggunakan StandardScaler. Tujuan dari proses ini adalah untuk menyamakan skala antar fitur agar algoritme tidak menjadi bias terhadap fitur dengan nilai besar seperti *bytes_in*, *bytes_out*, atau durasi koneksi. Standarisasi dilakukan terhadap semua

© 2025 The Author(s). Published by AIRA . This is an open access article under the CC BY-SA license (http://creativecommons.org/licenses/by-sa/4.0/)

e-ISSN 3063-4687

fitur numerik dalam data sampling menggunakan pendekatan transformasi *z-score*, sebagaimana telah digunakan dalam beberapa studi sebelumnya [6], [14], [17].

2.4 Penerapan Isolation Forest

Algoritme *Isolation Forest* digunakan untuk mendeteksi trafik anomali. Model ini dilatih menggunakan data yang telah distandarisasi dengan parameter `n_estimators=100`, `contamination=0.01`, `dan random_state=42`. Hasil prediksi diklasifikasikan menjadi dua kelas, yaitu 1 untuk trafik normal dan -1 untuk trafik anomali. Prediksi ini ditambahkan sebagai kolom baru *anomaly* dalam data.

Tabel 1. Parameter Model *Isolation Forest*

Parameter	Nilai
n_estimators	100
Contamination	0,01
random_state	42

Tahapan praproses pada uji sampling menggunakan *Isolation Forest* untuk evaluasi model deteksi anomali. Adapun parameter model *Isolation Forest*, diantaranya n_estimators, Contamination, dan random_state. Penggunaan contamination sebesar 0.01 didasarkan pada pendekatan umum di literatur ketika label *ground truth* terbatas [14], [18].

2.5 Visualisasi dan Evaluasi

Hasil deteksi divisualisasikan dalam bentuk scatter plot berdasarkan fitur *bytes_in* dan *bytes_out*, sebagaimana dilakukan dalam studi sebelumnya [9]. Evaluasi performa model dilakukan menggunakan *confusion matrix, precision, recall*, dan *F1-score* jika label *ground truth* tersedia [10], [12].

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dataset awal tersedia dalam format CSV dan terdiri atas 590.087 baris dengan 13 fitur awal yang merepresentasikan flow-level trafik jaringan. Sebelum digunakan untuk pelatihan model, dilakukan proses pembersihan (*cleaning*) dengan menghapus fitur-fitur yang tidak relevan dan redundan, sehingga diperoleh enam fitur utama yang menjadi fokus analisis. Fitur-fitur terpilih ditunjukkan pada Tabel 2 berikut:

Tabel 2. Fitur Dataset LUflow setelah *cleaning*

Nama Fitur	Tipe Data	Deskripsi Singkat
bytes_in	Numerik	Jumlah byte masuk pada sesi koneksi
bytes_out	Numerik	Jumlah byte keluar
entropy	Numerik	Tingkat acak trafik
duration	Numerik	Lama sesi (dalam satuan mikrodetik)
src_ip	Kategorikal	IP sumber (di-encode)
proto	Kategorikal	Protokol komunikasi (di-encode)

Meskipun dataset mencakup beberapa label trafik, analisis ini hanya difokuskan pada deteksi trafik outlier berbasis pola numerik tanpa membedakan tipe serangan tertentu. Hal ini

© 2025 The Author(s). Published by AIRA . This is an open access article under the CC BY-SA license (http://creativecommons.org/licenses/by-sa/4.0/)

e-ISSN 3063-4687

disebabkan oleh keterbatasan jumlah label bot yang tersedia serta dominasi trafik benign dalam data sampling, selanjutnya dilakukan praproses uji sampling dataset anomali.

3.2 Deteksi Anomali

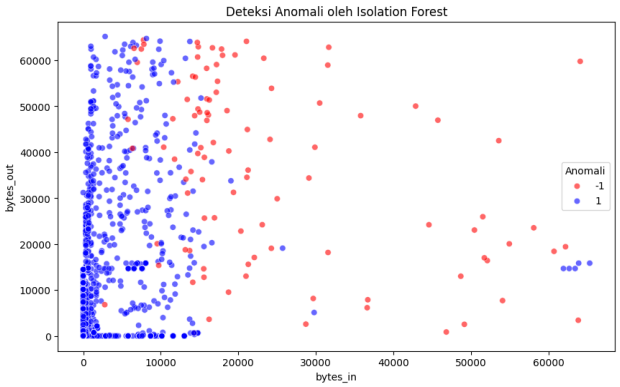
Model *Isolation Forest* dilatih menggunakan 10.000 baris data hasil preprocessing dan standarisasi. Dengan parameter `contamination=0.01`, model diasumsikan akan mendeteksi sekitar 1% data sebagai anomali. Hasil klasifikasi menunjukkan bahwa sebanyak **100 baris** terdeteksi sebagai trafik anomali (label = -1), sedangkan **9.900 baris** diklasifikasikan sebagai trafik normal (label = 1), sebagaimana ditunjukkan pada Gambar 3.

```
anomaly
 1    9900
-1     100
Name: count, dtype: int64
```

Gambar 3. Hasil uji Sampling Deteksi Anomali

Distribusi ini menunjukkan bahwa model bekerja sesuai dengan asumsi parameter awal. Trafik yang dikategorikan sebagai anomali umumnya memiliki nilai ekstrem pada fitur seperti *bytes_in*, *bytes_out*, dan *entropy*, yang mengindikasikan adanya penyimpangan signifikan terhadap karakteristik trafik mayoritas. Temuan ini menguatkan efektivitas pendekatan berbasis outlier dalam mengisolasi titik-titik data yang menyimpang secara statistik dari distribusi utama.

3.3 Visualisasi Hasil Deteksi



Gambar 4. Scatter plot hasil prediksi *Isolation Forest* antara fitur *bytes_in* dan *bytes_out*.

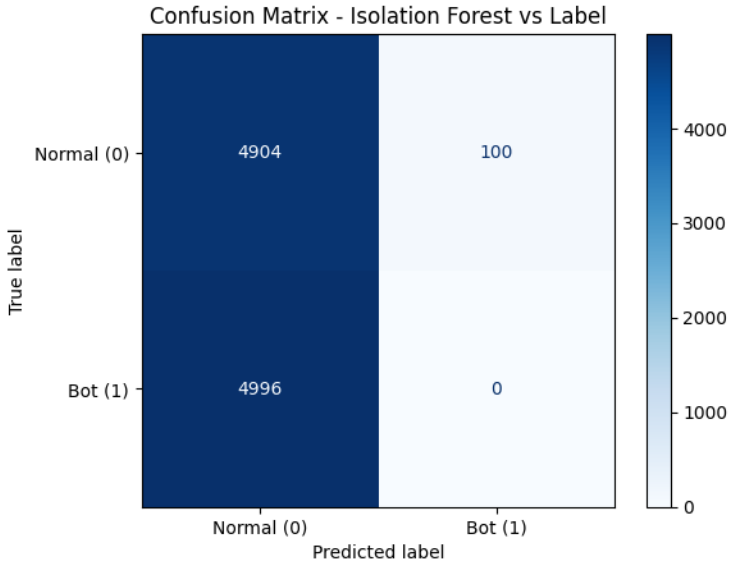
Gambar 4 menampilkan scatter plot antara bytes_in dan bytes_out dengan pewarnaan berdasarkan hasil deteksi model. Titik merah merepresentasikan trafik anomali (anomaly = -1), sedangkan titik biru merupakan trafik normal (anomaly = 1). Terlihat bahwa sebagian besar trafik normal membentuk klaster yang padat di daerah nilai rendah hingga menengah, sementara titik-titik anomali tersebar di luar distribusi umum, terutama pada nilai bytes yang sangat tinggi. Ini menunjukkan bahwa *Isolation Forest* berhasil mengidentifikasi trafik abnormal berdasarkan penyimpangan volume data.

3.4 Evaluasi Model

Evaluasi model dilakukan dengan membandingkan hasil prediksi terhadap label asli dari dataset (label). Hasil confusion matrix dan classification report dapat dilihat pada Gambar 2 dan Tabel 3. Model menunjukkan bahwa meskipun mampu mengenali beberapa trafik normal, tidak ada trafik yang dilabeli sebagai bot (1) berhasil diklasifikasikan dengan benar oleh model. Hal ini ditunjukkan dengan recall dan precision sebesar 0 untuk kelas bot. Dengan akurasi keseluruhan sebesar 49%, model cenderung hanya efektif dalam mengenali outlier statistik, bukan mendeteksi serangan bot secara eksplisit.

© 2025 The Author(s). Published by AIRA . This is an open access article under the CC BY-SA license (http://creativecommons.org/licenses/by-sa/4.0/)

e-ISSN 3063-4687



Gambar 5. *Confusion matrix* hasil prediksi *Isolation Forest* terhadap label *ground truth*. Matriks menunjukkan kelemahan model dalam mendeteksi kelas bot.

Sedangkan hasil tabel classification report menunjukkan bahwa nilai precision, recall, dan F1-score untuk kelas bot bernilai 0, yang menegaskan bahwa model tidak mampu mendeteksi serangan bot dengan baik.

Tabel 3. *Classification Report*

Class	Precision	Recall	F1-Score	Support
Normal (0)	0.4954	0.9800	0.6581	5004
Bot (1)	0.0000	0.0000	0.0000	4996
Accuracy			0.4904	10000

3.5 Analisis Fitur yang Mempengaruhi Deteksi Anomali

Meskipun *Isolation Forest* tidak menyediakan bobot fitur secara eksplisit, analisis distribusi dan korelasi fitur dapat memberikan gambaran kontribusi masing-masing terhadap deteksi anomali. Berdasarkan scatter plot (Gambar 1), fitur seperti bytes_out dan entropy memperlihatkan batas distribusi yang jelas antara trafik normal dan trafik yang diklasifikasikan sebagai anomali. Hal ini menunjukkan bahwa pola volume data keluar serta tingkat ketidakteraturan (*entropy*) memiliki dampak signifikan terhadap proses pemisahan antara data mayoritas (*inlier*) dan data menyimpang (*outlier*).

Temuan ini sejalan dengan penelitian Liu et al., yang menunjukkan bahwa lonjakan data keluar dan entropi tinggi sering menjadi indikator kuat dalam deteksi anomali pada data trafik jaringan. Selain itu, pendekatan yang dimodifikasi dari *Isolation Forest* seperti *Extended Isolation Forest* (EIF) memungkinkan analisis kontribusi fitur terhadap proses isolasi anomali, dan fitur volume data serta entropi terbukti dominan dalam mengidentifikasi trafik menyimpang.

3.6 Analisis Fitur yang Mempengaruhi Deteksi Anomali

Kelebihan:

1. Tidak memerlukan label dalam proses pelatihan (*unsupervised*), sehingga cocok digunakan untuk data jaringan yang tidak sepenuhnya terlabel.
2. Memiliki kompleksitas waktu yang rendah (*O(n log n)*) dan dapat diimplementasikan secara efisien pada dataset besar maupun dalam platform cloud seperti Google Colab.

© 2025 The Author(s). Published by AIRA . This is an open access article under the CC BY-SA license (http://creativecommons.org/licenses/by-sa/4.0/)

3. Mampu mendeteksi anomali ekstrem (*outlier*) yang secara statistik sangat berbeda dari distribusi umum.
4. Kekurangan:

5. Tidak cukup sensitif terhadap trafik bot yang menyamar sebagai trafik normal karena model hanya mengenali anomali yang sangat menyimpang secara statistik.
6. Ketergantungan terhadap parameter contamination yang perlu disesuaikan dengan karakteristik data target.
7. Rentan menghasilkan false positive ketika titik borderline diklasifikasikan sebagai anomali.
8. Interpretabilitas model secara default terbatas, meskipun varian seperti EIF dan ExIFFI telah dikembangkan untuk meningkatkan transparansi deteksi.

4. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritme *Isolation Forest* efektif dalam mendeteksi trafik anomali berbasis pola numerik dari data jaringan. Dengan memanfaatkan dataset LUFlow, model mampu mengidentifikasi trafik menyimpang yang secara statistik menunjukkan perilaku tidak lazim, seperti volume data yang tinggi dan nilai entropi yang ekstrem. Visualisasi hasil dalam bentuk scatter plot membantu menggambarkan secara eksplisit distribusi antara trafik normal dan anomali. Evaluasi kuantitatif menggunakan confusion matrix dan classification report mengungkapkan bahwa meskipun model memiliki akurasi dalam mengidentifikasi anomali, terdapat keterbatasan dalam mengenali jenis serangan spesifik seperti bot, terutama ketika data tidak dilabeli secara merata. Mengingat karakteristik *Isolation Forest* yang berbasis unsupervised, pendekatan ini sangat sesuai untuk lingkungan jaringan dengan ketersediaan label terbatas. Akan tetapi, untuk mendeteksi serangan yang lebih kompleks dan tersembunyi seperti botnet atau advanced persistent threats (APT), dibutuhkan integrasi dengan teknik lain seperti pembelajaran berbasis representasi (representation learning), autoencoder, atau pendekatan berbasis graf untuk menangkap hubungan antar flow. Untuk pengembangan lebih lanjut, penelitian serupa sebaiknya mempertimbangkan skenario data streaming guna mengevaluasi performa model dalam kondisi real-time, serta mengakomodasi perubahan pola trafik (concept drift) yang terjadi secara dinamis. Eksplorasi terhadap teknik ensemble yang menggabungkan beberapa metode deteksi juga patut dipertimbangkan untuk meningkatkan sensitivitas terhadap berbagai jenis anomali. Selain itu, integrasi model deteksi ini ke dalam sistem monitoring jaringan secara aktual dapat menjadi langkah konkret dalam menguji efektivitasnya dalam konteks operasional. Perlu juga dilakukan penelitian lanjutan mengenai interpretabilitas model, agar hasil deteksi dapat dijelaskan secara lebih transparan kepada analis keamanan siber. Hal ini penting terutama dalam konteks pengambilan keputusan yang memerlukan justifikasi terhadap deteksi yang dilakukan oleh sistem. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi terhadap pemahaman teknis penggunaan *Isolation Forest*, tetapi juga membuka peluang besar untuk pengembangan sistem deteksi anomali yang lebih adaptif dan aplikatif di masa mendatang.

DAFTAR PUSTAKA

[1] Y. Feng *et al.*, “An improved X-means and isolation forest based methodology for network traffic anomaly detection,” *PLoS One*, vol. 17, no. 1, p. e0263423, Jan. 2022, doi: 10.1371/journal.pone.0263423.

[2] R. Ardiansyah, L. Sunardi, and Martadinata A, “IMPLEMENTASI METODE ISOLATION FOREST UNTUK DETEKSI ANOMALI DALAM DATA JARINGAN,” *Universitas Bina Insan Lubuklinggau*, vol. 4, pp. 208–216, 2025, Accessed: Jun. 13, 2025. [Online]. Available: https://semnas.univbinainsan.ac.id/index.php/escaf/article/view/851

© 2025 The Author(s). Published by AIRA . This is an open access article under the CC BY-SA license (http://creativecommons.org/licenses/by-sa/4.0/)

e-ISSN 3063-4687

[3] A. Kharitonov, A. Nahhas, M. Pohl, and K. Turowski, "Comparative analysis of machine learning models for anomaly detection in manufacturing," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 1288–1297. doi: 10.1016/j.procs.2022.01.330.

[4] Milka Wijayanti Sunarto, Dendy Kurniawan, Edy Siswanto, and Haris Ihsanil Huda, "Deteksi Anomali Menggunakan Extended Isolation Forest (Eif)," *Teknik: Jurnal Ilmu Teknik dan Informatika*, vol. 1, no. 2, pp. 96–111, May 2023, doi: 10.51903/teknik.v1i2.324.

[5] M. V. Mahoney, "Network traffic anomaly detection based on packet bytes," in *Proceedings of the 2003 ACM symposium on Applied computing*, New York, NY, USA: ACM, Mar. 2003, pp. 346–350. doi: 10.1145/952532.952601.

[6] M. Mutmainah and W. Yustanti, "Studi Komparasi Local Outlier Factor (LOF) dan Isolation Forest (IF) pada Analisis Anomali Kinerja Dosen," *Journal of Informatics and Computer Science (JINACS)*, vol. 6, no. 02, pp. 532–540, Jul. 2024, doi: 10.26740/jinacs.v6n02.p532-540.

[7] R. Alsini, A. Almakrab, A. Ibrahim, and X. Ma, "Improving the outlier detection method in concrete mix design by combining the isolation forest and local outlier factor," *Constr Build Mater*, vol. 270, p. 121396, Feb. 2021, doi: 10.1016/j.conbuildmat.2020.121396.

[8] N. Al Khater and R. E. Overill, "Network traffic classification techniques and challenges," in *2015 Tenth International Conference on Digital Information Management (ICDIM)*, IEEE, Oct. 2015, pp. 43–48. doi: 10.1109/ICDIM.2015.7381869.

[9] G. Nychis, V. Sekar, D. G. Andersen, H. Kim, and H. Zhang, "An empirical evaluation of entropy-based traffic anomaly detection," in *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, New York, NY, USA: ACM, Oct. 2008, pp. 151–156. doi: 10.1145/1452520.1452539.

[10] A. S. Shukla and R. Maurya, "Entropy-Based Anomaly Detection in a Network," *Wirel Pers Commun*, vol. 99, no. 4, pp. 1487–1501, Apr. 2018, doi: 10.1007/s11277-018-5288-2.

[11] S. A. Elsaid and A. Binbusayyis, "An optimized isolation forest based intrusion detection system for heterogeneous and streaming data in the industrial Internet of Things (IIoT) networks," *Discover Applied Sciences*, vol. 6, no. 9, p. 483, Sep. 2024, doi: 10.1007/s42452-024-06165-w.

[12] Z. Ding and M. Fei, "An Anomaly Detection Approach Based on Isolation Forest Algorithm for Streaming Data using Sliding Window," *IFAC Proceedings Volumes*, vol. 46, no. 20, pp. 12–17, 2013, doi: 10.3182/20130902-3-CN-3020.00044.

[13] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, "A survey of deep learning-based network anomaly detection," *Cluster Comput*, vol. 22, no. S1, pp. 949–961, Jan. 2019, doi: 10.1007/s10586-017-1117-8.

[14] 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.

[15] X. Chun-Hui, S. Chen, B. Cong-Xiao, and L. Xing, "Anomaly Detection in Network Management System Based on Isolation Forest," in *2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC)*, IEEE, Apr. 2018, pp. 56–60. doi: 10.1109/ICNISC.2018.00019.

[16] M. U. Togbe *et al.*, "Anomaly Detection for Data Streams Based on Isolation Forest Using Scikit-Multiflow," 2020, pp. 15–30. doi: 10.1007/978-3-030-58811-3_2.

[17] Z. Yang *et al.*, "A systematic literature review of methods and datasets for anomaly-based network intrusion detection," *Comput Secur*, vol. 116, p. 102675, May 2022, doi: 10.1016/j.cose.2022.102675.

[18] W. Wu, J. Alvarez, C. Liu, and H.-M. Sun, "Bot detection using unsupervised machine learning," *Microsystem Technologies*, vol. 24, no. 1, pp. 209–217, Jan. 2018, doi: 10.1007/s00542-016-3237-0.

[19] LUFlow Network Intrusion Detection Data Set, https://www.kaggle.com/datasets/mryanm/luflow-network-intrusion-detection-data-set?resource=download&select=2022

© 2025 The Author(s). Published by AIRA . This is an open access article under the CC BY-SA license (http://creativecommons.org/licenses/by-sa/4.0/)