

ANALISIS DATA KEUANGAN MENGGUNAKAN ISOLATION FOREST UNTUK MENINGKATKAN AKURASI DETEKSI ANOMALI

FINANCIAL DATA ANALYSIS USING ISOLATION FOREST TO IMPROVE ANOMALY DETECTION ACCURACY

Gladys Anawai Ohyver¹, Jap Tji Beng^{1*}, Dedi Trisnawarman¹, Sri Tiatri², Daniel Adolf Ohyver³

Fakultas Teknologi Informasi, Program Studi Sarjana Sistem Informasi, Universitas Tarumanagara,
Jakarta¹

Fakultas Psikologi, Program Studi Magister Psikologi, Universitas Tarumanagara, Jakarta²
Hospitality, Politeknik Pariwisata Makassar³

t.jap@untar.ac.id^{1*}

ABSTRACT

Financial data analysis plays an important role in ensuring efficiency, transparency, and accuracy in budget management. A common issue often encountered is the presence of irregular data, which may indicate recording errors, inefficiency, or potential fraud. This study aims to apply the Isolation Forest method, an unsupervised ensemble learning algorithm, for the automatic detection of anomalies in financial data. The method works by isolating data points that deviate from general patterns through the construction of random decision trees (random partitioning). The fewer steps required to isolate a data point, the higher its likelihood of being an anomaly. The results show that Isolation Forest effectively identifies irregularities with high efficiency while providing an objective representation of financial data patterns. In conclusion, this method demonstrates potential as a reliable and adaptive analytical approach to support continuous auditing and financial monitoring processes.

Keywords: Isolation Forest, Anomaly Detection, Financial Data, Data Analysis

ABSTRAK

Analisis data keuangan merupakan langkah penting dalam memastikan efisiensi, transparansi, dan keakuratan pengelolaan anggaran. Permasalahan umum yang sering muncul adalah adanya data yang menyimpang yang dapat mengindikasikan kesalahan pencatatan, inefisiensi, atau potensi kecurangan. Penelitian ini bertujuan menerapkan metode *Isolation Forest*, yaitu algoritma *unsupervised learning* berbasis *ensemble*, untuk mendeteksi anomali pada data keuangan secara otomatis. Metode ini bekerja dengan mengisolasi data yang berbeda dari pola umum melalui pembentukan pohon keputusan acak atau *random partitioning*. Semakin sedikit langkah yang dibutuhkan untuk mengisolasi suatu data, semakin besar kemungkinan bahwa data tersebut merupakan anomali. Hasil penelitian menunjukkan bahwa *Isolation Forest* efektif dalam mengidentifikasi penyimpangan dengan efisiensi tinggi dan memberikan gambaran yang objektif terhadap pola data keuangan. Kesimpulannya, metode ini berpotensi menjadi pendekatan analitik yang adaptif dan andal untuk mendukung audit serta pengawasan keuangan secara berkelanjutan.

Kata Kunci: Isolation Forest, Deteksi Anomali, Data Keuangan, Analisis Data.

PENDAHULUAN

Dalam pengelolaan keuangan modern, data berperan sebagai sumber utama untuk pengambilan keputusan strategis (Bernard, 2021; Kimball & Ross, 2013). Namun, volume data keuangan yang besar dan kompleks sering kali mengandung nilai-nilai yang menyimpang dari pola umum yang dapat mengindikasikan kesalahan pencatatan atau potensi kecurangan (Zulfikar et al., 2023). Penyimpangan ini dapat muncul akibat kesalahan manusia, sistem pencatatan yang

tidak konsisten, ataupun indikasi *fraud*. Oleh karena itu, dibutuhkan pendekatan analitik yang mampu mengenali pola anomali secara otomatis dan akurat tanpa ketergantungan pada pengawasan manual.

Sejumlah penelitian terdahulu memperlihatkan bahwa metode *machine learning* memiliki kemampuan adaptif dalam mendeteksi anomali keuangan dengan akurasi tinggi. Hilal et al., (2022) menunjukkan bahwa algoritma *unsupervised learning* seperti *Isolation Forest* dapat mengidentifikasi indikasi

fraud secara lebih efektif dibandingkan pendekatan statistik tradisional. *Isolation Forest* sendiri sendiri pertama kali diperkenalkan oleh Liu, Ting, dan Zhou (2008) dan telah menjadi salah satu metode deteksi anomali paling banyak digunakan karena sifatnya yang efisien serta tidak bergantung pada asumsi distribusi data. Studi lanjutan seperti Xu et al., (2023) menguatkan efektivitas konsep isolasi dalam mendeteksi penyimpangan pada dataset berskala besar dan berdimensi tinggi. Selain itu, penelitian dari Bakumenko & Elragal (2022) membuktikan relevansi metode ini dalam menganalisis data keuangan, khususnya ketika variabel bersifat numerik dan memiliki kecenderungan menghasilkan *outlier*.

Penelitian ini bertujuan untuk menganalisis penerapan *Isolation Forest* dalam mendeteksi anomali pada data keuangan, dengan fokus pada kemampuan model dalam mengidentifikasi transaksi atau nilai yang tidak wajar. Selain itu, penelitian ini memperluas konteks dengan meninjau penerapan analitik keuangan modern yang menggabungkan pendekatan data-driven decision making untuk mendukung audit dan efisiensi organisasi (Park, 2022). Melalui pendekatan ini, diharapkan diperoleh hasil analisis yang mampu memperkuat keandalan Audit internal serta meningkatkan transparansi dan efisiensi pengelolaan anggaran keuangan.

METODE

Penelitian ini merupakan penelitian terapan (*applied research*) dengan pendekatan kuantitatif yang berfokus pada penerapan metode *Isolation Forest* dalam analisis data keuangan untuk mendeteksi anomali (Patel, 2019). Penelitian ini dilakukan dengan menggunakan bahasa pemrograman Python sebagai alat utama analisis dengan pustaka *pandas*, *NumPy*, dan *scikit-learn* (VanderPlas, 2023), yang memungkinkan proses pra-pemrosesan,

normalisasi, dan evaluasi dilakukan secara sistematis dan komprehensif.

Pengumpulan Data

Data keuangan yang digunakan dalam penelitian ini terdiri atas variabel nominal seperti total anggaran, nilai realisasi, serta perhitungan selisih dan rasio antara keduanya. Data bersumber dari laporan transaksi keuangan dengan periode tertentu yang sudah terdokumentasi (Hilal et al., 2022). Seluruh data dianalisis menggunakan teknik total sampling karena karakteristik penelitian *data mining* yang membutuhkan keseluruhan dataset untuk mempelajari pola (VanderPlas, 2023).

Pra-pemrosesan Data

Tahap ini bertujuan untuk memastikan data dalam kondisi bersih, terstruktur, dan siap digunakan dalam proses analitik (Kimball & Ross, 2013). Proses pra-pemrosesan dilakukan menggunakan bahasa pemrograman Python dengan dukungan pustaka *pandas* dan *NumPy* untuk transformasi dan pembersihan data, serta *scikit-learn* digunakan untuk proses scaling numerik dan penerapan model *Isolation Forest*. Langkah-langkah pra-pemrosesan mencakup:

1. Data Cleaning

Menghapus nilai kosong, data duplikat, serta memperbaiki kesalahan input seperti nilai negatif atau format angka yang tidak konsisten.

2. Normalisasi Data

Mengubah skala data numerik menggunakan metode Min-Max Scaling agar semua variabel berada pada rentang yang sama (0–1). Langkah ini penting untuk menjaga konsistensi performa algoritma *Isolation Forest* yang sensitif terhadap skala data.

3. Feature Selection

Memilih atribut yang relevan untuk proses analisis, seperti total anggaran, realisasi, dan rasio penggunaan anggaran. Atribut-atribut yang tidak

berpengaruh terhadap deteksi anomali dihilangkan.

4. Pembuatan Variabel Turunan

Tahap ini bertujuan menambah fitur-fitur baru yang dapat memperkaya konteks analisis dan membantu model dalam mengenali pola penyimpangan pada data keuangan. Dua variabel turunan utama yang digunakan adalah Selisih dan Rasio. Variabel Selisih diperoleh dari perbedaan antara nilai realisasi dan anggaran. Sedangkan Rasio diperoleh dari pembagian antara realisasi dan anggaran. Nilai Selisih menggambarkan deviasi antara anggaran dan realisasi, sedangkan Rasio menunjukkan tingkat efisiensi penggunaan anggaran. Kedua variabel ini memberikan informasi tambahan yang relevan bagi model *Isolation Forest* dalam mengidentifikasi transaksi dengan karakteristik penyimpangan dari pola umum (Bakumenko & Elragal, 2022).

Exploratory Data Analysis (EDA)

Sebelum penerapan model, dilakukan analisis eksploratif untuk memahami distribusi, tren, dan korelasi antarvariabel. Tahapan ini mencakup:

1. Visualisasi distribusi nilai anggaran dan realisasi menggunakan boxplot dan histogram untuk mengidentifikasi nilai ekstrem secara visual.
2. Analisis korelasi antarvariabel guna mengetahui hubungan yang signifikan antara rasio penggunaan dana dan total pengeluaran.
3. Identifikasi awal indikasi anomali melalui grafik scatter dua dimensi sebelum dilakukan pemodelan dengan *Isolation Forest*.

Algoritma Isolation Forest

Metode *Isolation Forest* digunakan untuk mendeteksi anomali berdasarkan prinsip isolation, yaitu memisahkan setiap titik data melalui pembentukan sejumlah pohon keputusan acak. Semakin sedikit langkah yang dibutuhkan untuk

mengisolasi suatu titik, semakin tinggi tingkat anomali data tersebut. Skor anomali $s(x, n)$ dihitung dengan **Persamaan (1)**:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (1)$$

dengan keterangan:

- $s(x, n)$ = *anomaly score* untuk titik data x
- $E(h(x))$ = panjang rata-rata path untuk mengisolasi x di seluruh pohon
- n = jumlah sampel dalam dataset
- $c(n)$ = fungsi normalisasi untuk rata-rata panjang path pohon biner.

Fungsi normalisasi $c(n)$ untuk menghitung panjang *path* dirumuskan sebagai **Persamaan (2)**:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \quad (2)$$

di mana $H(i) = \ln(i) + 0.5772$ merupakan bilangan harmonik yang digunakan untuk menyesuaikan skala panjang *path* terhadap ukuran sampel. Dengan demikian, fungsi $c(n)$ berperan menjaga konsistensi perhitungan skor anomali antar dataset dengan ukuran berbeda.

Model diimplementasikan dengan parameter utama $n_estimators = 100$ dan $contamination = 0.05$, sesuai proporsi data yang diduga anomali. Nilai skor anomali $s(x, n)$ berada pada rentang 0 hingga 1. Semakin mendekati 1, data tersebut semakin mudah diisolasi dan cenderung merupakan anomali. Namun, skor mendekati 0 menunjukkan data normal atau konsisten dengan pola umum.

Evaluasi

Tahap evaluasi dilakukan dengan meninjau pola hasil deteksi dan validitas anomali yang ditemukan. Karena penelitian ini bersifat unsupervised, evaluasi dilakukan melalui:

1. Analisis Visual dan Logis

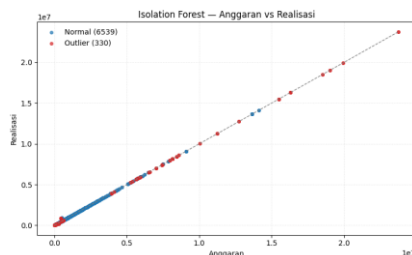
Menilai apakah data yang diklasifikasikan sebagai anomali benar-benar menunjukkan penyimpangan

- signifikan, misalnya rasio realisasi jauh di atas 100% atau jauh di bawah target.
2. Perbandingan Nilai Skor
Mengelompokkan hasil berdasarkan rentang *anomaly score* untuk menilai intensitas penyimpangan. Contoh kategori:
 - a) Skor 0.5–0.7 → anomali ringan
 - b) Skor 0.7–0.9 → anomali sedang
 - c) Skor >0.9 → anomali kuat
 3. Validasi Domain *Knowledges*
Hasil deteksi dikaji berdasarkan konteks keuangan untuk memastikan bahwa anomali yang terdeteksi memang relevan secara praktis.

HASIL DAN PEMBAHASAN

Hasil analisis data keuangan menggunakan metode *Isolation Forest* menunjukkan kemampuan model dalam mengidentifikasi transaksi yang menyimpang dari pola umum hubungan antara nilai anggaran dan realisasi. Hasil pemodelan menunjukkan bahwa dari total 6.869 data keuangan, sebanyak 6.539 data (95,2%) terdeteksi sebagai normal, sementara 330 data (4,8%) terindikasi sebagai anomali.

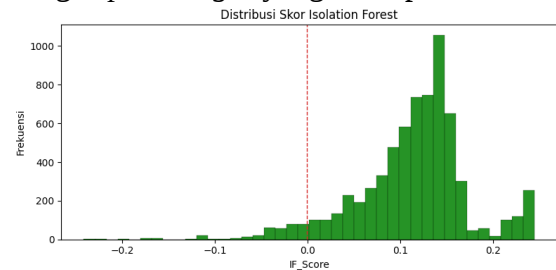
Gambar 1 memperlihatkan sebaran data pada grafik “Anggaran vs Realisasi”, di mana titik biru merepresentasikan data normal dan titik merah menunjukkan data yang diklasifikasikan sebagai anomali. Pola umum memperlihatkan hubungan linier positif antara anggaran dan realisasi, menandakan sebagian besar transaksi berjalan proporsional. Namun, beberapa titik merah yang menyimpang di atas maupun di bawah garis tren menunjukkan transaksi dengan rasio realisasi yang tidak wajar.



Gambar 1. Sebaran Data Anggaran dan Realisasi

Sumber : Dokumentasi Pribadi

Gambar 2 menampilkan histogram distribusi skor *Isolation Forest* (IF_Score), yang menggambarkan tingkat keanehan setiap data. Histogram skor *Isolation Forest* mengindikasikan bahwa sebagian besar data memiliki rentang skor antara 0,05–0,15. Skor yang lebih rendah menunjukkan kemungkinan anomali yang lebih kuat. Hasil ini mendukung temuan penelitian sebelumnya seperti (Xu et al., 2023) dan (Zulfikar et al., 2023), yang menunjukkan bahwa penyimpangan yang signifikan pada selisih atau rasio realisasi dapat terisolasi dengan *path length* yang lebih pendek.



Gambar 2. Distribusi Skor Isolation Forest

Sumber : Dokumentasi Pribadi

Selain itu, **Gambar 3** menampilkan sepuluh data dengan nilai IF_Score terendah yang diidentifikasi sebagai *top outliers*. Sebagian besar data anomali tersebut memiliki nilai selisih negatif yang besar dan rasio realisasi yang rendah, seperti 0,41 atau 0,85 yang menandakan ketidaksesuaian antara alokasi dan pelaksanaan kegiatan.

sk_item	sk_prodi	sk_time	anggaran	realisasi	unit_harga	volume	selisih	rasio	IF_Score	IF_Status
1172	2	76	6000	5100	5100	1	-900	0.850000	-0.241175	Outlier
1172	2	537	6000	5100	5100	1	-900	0.850000	-0.241175	Outlier
730	4	168	290000	120412	60206	2	-169588	0.415214	-0.219221	Outlier
730	4	526	290000	120412	60206	2	-169588	0.415214	-0.219221	Outlier
755	2	465	225000	220100	220100	1	-4900	0.978222	-0.197577	Outlier
756	2	465	225000	220100	220100	1	-4900	0.978222	-0.197577	Outlier
270	2	465	10000	9300	9300	1	-700	0.930000	-0.196223	Outlier
270	2	466	10000	9300	9300	1	-700	0.930000	-0.196223	Outlier
221	4	295	5000	4350	4350	1	-650	0.870000	-0.175798	Outlier
222	4	295	5000	4350	4350	1	-650	0.870000	-0.175798	Outlier

Gambar 3. Data dengan Skor Anomali Terendah

Sumber : Dokumentasi Pribadi

Berdasarkan hasil pemodelan, metode *Isolation Forest* terbukti efektif dalam memisahkan data normal dan anomali pada konteks keuangan tanpa memerlukan label atau asumsi distribusi data tertentu. Hubungan linier antara variabel anggaran dan realisasi dimanfaatkan oleh model untuk mengenali pola umum transaksi normal. Ketika ditemukan nilai realisasi yang terlalu tinggi

atau terlalu rendah terhadap anggaran, model secara otomatis mengisolasi data tersebut sebagai anomali karena pola partisipasinya berbeda signifikan dibandingkan data normal.

Temuan ini menunjukkan bahwa *Isolation Forest* mampu mendeteksi dua kategori penyimpangan utama, yaitu:

a) *Overbudget Anomaly*

Merupakan realisasi yang melebihi alokasi anggaran yang telah ditetapkan. Kasus ini berpotensi mengindikasikan pembengkakan biaya, kesalahan dalam pencatatan transaksi, atau penggunaan dana yang tidak efisien.

b) *Underbudget Anomaly*

Merupakan nilai realisasi yang jauh di bawah anggaran. Meskipun secara kasat mata terlihat positif, model tetap mengidentifikasinya sebagai anomali karena distribusi nilai tersebut berbeda dari pola umum. Hal ini dapat terjadi akibat kegiatan yang belum terealisasi sepenuhnya, efisiensi ekstrem, atau perencanaan anggaran yang tidak akurat.

Secara analitik, *Isolation Forest* memberikan hasil yang stabil dan efisien pada dataset besar lebih dari 6.000 baris, dengan waktu komputasi singkat dan tingkat deteksi yang konsisten. Hasil visualisasi juga memperlihatkan pemisahan yang jelas antara data normal dan anomali yang memperkuat validitas model dalam mengidentifikasi pola penyimpangan keuangan.

Secara konseptual, hasil penelitian ini menegaskan bahwa pendekatan *machine learning* berbasis isolasi seperti *Isolation Forest* dapat menjadi alternatif yang lebih adaptif dibandingkan metode statistik tradisional dalam menganalisis data keuangan yang kompleks dan tidak selalu berdistribusi normal. Implementasinya dapat mendukung proses audit internal serta pengawasan anggaran secara otomatis dan membantu pengambil keputusan untuk segera menindaklanjuti transaksi yang berpotensi bermasalah, baik yang bersifat pemborosan maupun realisasi yang terlalu rendah.

SIMPULAN

1. Penelitian ini berhasil menerapkan metode *Isolation Forest* untuk mendeteksi anomali pada data keuangan secara otomatis dan efektif. Model mampu memisahkan data normal dan anomali berdasarkan pola hubungan antara anggaran dan realisasi tanpa memerlukan label data atau asumsi distribusi tertentu.
2. Hasil analisis menunjukkan bahwa dari total 6.869 data keuangan, sebanyak 330 data (sekitar 4,8%) terdeteksi sebagai anomali. Sebagian besar anomali memiliki selisih negatif yang besar dan rasio realisasi yang terlalu rendah atau terlalu tinggi, menandakan adanya ketidaksesuaian antara alokasi dan realisasi anggaran.
3. Visualisasi hasil berupa sebaran data dan distribusi skor *Isolation Forest* menunjukkan bahwa metode ini mampu mengidentifikasi penyimpangan dengan akurasi dan efisiensi tinggi. Nilai skor anomali yang rendah menunjukkan tingkat keanehan data yang signifikan terhadap pola umum.
4. Secara keseluruhan, *Isolation Forest* terbukti menjadi pendekatan analitik yang adaptif dan dapat diandalkan dalam mendukung proses audit dan pengawasan keuangan berbasis data. Hasil penelitian ini dapat menjadi dasar untuk pengembangan sistem deteksi anomali keuangan yang lebih cerdas, otomatis, dan berkelanjutan di masa depan.
5. Berdasarkan hasil tersebut, organisasi dapat mengintegrasikan model ini ke dalam *dashboard Business Intelligence* untuk menyediakan sistem peringatan (*early warning system*), selaras dengan rekomendasi penelitian sebelumnya (Lawrence et al., 2025; Putri et al., 2024).
6. Penelitian selanjutnya disarankan menambahkan variabel tambahan seperti kategori transaksi atau waktu untuk meningkatkan akurasi model serta

melakukan perbandingan metode *anomaly detection* lainnya.

DAFTAR PUSTAKA

- Bakumenko, A., & Elragal, A. (2022). Detecting Anomalies in Financial Data Using Machine Learning Algorithms. *Systems*, 10(5). <https://doi.org/10.3390/systems10050130>
- Bernard, M. (2021). *Data strategy: How to profit from a world of big data, analytics, and the Internet of Things*. Kogan Page.
- Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Expert Systems with Applications*, 193, 116429. <https://doi.org/https://doi.org/10.1016/j.eswa.2021.116429>
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (Third). John Wiley & Sons, Inc.
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. *2008 Eighth IEEE International Conference on Data Mining*, 413–422. <https://doi.org/10.1109/ICDM.2008.17>
- Park, C. (2022). A Comparative Study for Outlier Detection Methods in High Dimensional Text Data. *Journal of Artificial Intelligence and Soft Computing Research*, 13, 5–17. <https://doi.org/10.2478/jaiscr-2023-0001>
- Patel, A. A. . (2019). *Hands-on unsupervised learning using Python : how to build applied machine learning solutions from unlabeled data*. O'Reilly Media.
- VanderPlas, J. (2023). *Python data science handbook: Essential tools for working with data* (2nd ed.). O'Reilly Media.
- Xu, H., Pang, G., Wang, Y., & Wang, Y. (2023). Deep Isolation Forest for Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering*, PP, 1–14. <https://doi.org/10.1109/TKDE.2023.3270293>
- Zulfikar, A., Rahmani, F. A., & Azizah, N. (2023). Deteksi Anomali Menggunakan Isolation Forest Belanja Barang Persediaan Konsumsi Pada Satuan Kerja Kepolisian Republik Indonesia. *Jurnal Manajemen Perbendaharaan*, 4(1), 1–15. <https://doi.org/10.33105/jmp.v4i1.435>
- Lawrence, V., Beng, J. T., Wasino, W., Tiatri, S., Nurkholiza, R., Salsabila, T. M., & Sefira, F. M. (2025). Design of a data warehouse for customer segmentation prediction at PT X. *INTECOMS: Journal of Information Technology and Computer Science*, 8(5). <https://doi.org/10.31539/he6dx426>
- Limbor, E. G., Beng, J. T., Arisandi, D., Tiatri, S., Nurkholiza, R., Salsabila, T. M., & Dinatha, V. O. D. (2025). Designing a data mart for optimizing clothing sales at C Store. *INTECOMS: Journal of Information Technology and Computer Science*, 8(5). <https://doi.org/10.31539/7k3f3443>
- Putri, T. A., Angela, O., Ramdhani Qadriah, S. A., Trisnawarman, D., & Beng, J. T. (2024). Pengembangan Early Warning System dengan SmtClient untuk meningkatkan kinerja manajemen data akreditasi di PTS Y. *INTECOMS: Journal of Information Technology and Computer Science*, 7(6). <https://doi.org/10.31539/intecom.v7i6.12876>
- Ruben, R., & Trisnawarman, D. (2024). Desain dashboard untuk analisis harga pangan di Indonesia. *INTECOMS: Journal of Information Technology and Computer Science*, 7(3).

- <https://doi.org/10.31539/intecom.v7i3.10402>
- Mok, J. A., & Trisnawarman, D. (2024). Perancangan data mart penjualan pada PT XYZ. *INTECOMS: Journal of Information Technology and Computer Science*, 7(6). <https://doi.org/10.31539/intecom.v7i6.12960>
- Ohyver, D. A., Hanafi, H., & Muhtasom, A. (2023). Homestay digital marketing in tourism villages in South Sulawesi (Case study in 3 featured tourism villages). *Return: Study of Management, Economic and Business*, 2(10), 1035–1051.
- Masri, M., Judijanto, L., Harto, B., Ohyver, D. A., Kusumastuti, S. Y., & Rukmana, A. Y. (2025). *AI dan ekonomi berbasis data: Solusi inovatif untuk pengambilan keputusan ekonomi oleh entrepreneur*. Jambi: Sonpedia Publishing. ISBN 978-623-514-745-1.
- Airlangga, G. (2024, April). Advanced machine learning techniques for seismic anomaly detection in Indonesia: A comparative study of LOF, Isolation Forest, and One-Class SVM. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika & Statistika*, 5(1), 49-61. <https://doi.org/10.46306/lb.v5i1.490>
- Dewa, R. P., & Windarto, W. (2024). Deteksi anomali jaringan menggunakan Isolation Forest pada log Wazuh dengan pemberitahuan WhatsApp di PT XYZ. *Kresna: Jurnal Riset dan Pengabdian Masyarakat*, 4(2), 208-216. <https://doi.org/10.36080/kresna.v4i2.170>
- Yuliano, A. P. (2025). Deteksi trafik anomali berdasarkan pola trafik menggunakan Isolation Forest. *COSMIC: Jurnal Ilmiah Aira*