

Analisis Komparatif Model Unsupervised Learning (Isolation Forest Vs. Autoencoder) Untuk Deteksi Anomali Pada Klaim Asuransi Kesehatan

Muhammad Rosyadi¹, Susandri², Nurul Nuzilah Lestari³, Ahmad Zamsuri⁴

^{1,2,3} Program Studi Magister Ilmu Komputer, Sekolah Pasca Sarjana, Universitas Lancang Kuning, Indonesia

^{1,2,3} Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015

e-mail: ¹muhammadrosyadi26@gmail.com, ³susandri@unilak.ac.id ,

³nurulnuzilah01@gmail.com, ⁴ahmadzamsuri@unilak.ac.id

Abstrak

Asuransi merupakan produk keuangan yang berfungsi melindungi nasabah dari potensi risiko di masa depan yang tak terduga. Salah satu produk utamanya adalah asuransi kesehatan, yang sangat membantu untuk menutupi biaya pengobatan yang mahal, terutama bagi orang-orang yang tidak memiliki sumber daya finansial yang cukup. Deteksi anomali pada klaim asuransi kesehatan merupakan langkah krusial dalam mencegah kecurangan dan menjaga efisiensi operasional perusahaan asuransi. Penelitian ini membandingkan dua model unsupervised Learning, yaitu Isolation Forest dan Autoencoder, dalam mendeteksi anomali pada data klaim asuransi kesehatan. Evaluasi dilakukan menggunakan metrik AUC, akurasi, precision, recall, dan F1-score. Hasil menunjukkan bahwa model Isolation Forest memiliki performa yang lebih stabil dan unggul dalam hal akurasi dan F1-score dibandingkan Autoencoder. Sementara itu, Autoencoder menunjukkan nilai recall yang lebih tinggi dalam mendeteksi klaim anomali, namun precision yang rendah menurunkan efektivitas deteksi secara keseluruhan. Berdasarkan hasil tersebut, Isolation Forest direkomendasikan sebagai metode yang lebih andal untuk deteksi anomali pada kasus ini.

Kata Kunci: Unsupervised Learning, Isolation Forest, Autoencoder, Deteksi Anomali, Klaim Asuransi Kesehatan.

Abstract

Insurance is a financial product that serves to protect customers from unexpected future risks. One of its main products is health insurance, which plays a vital role in covering high medical expenses, especially for individuals who lack sufficient financial resources. Anomaly detection in health insurance claims is a crucial step in preventing fraud and maintaining the operational efficiency of insurance companies. This study compares two unsupervised Learning models, namely Isolation Forest and Autoencoder, in detecting anomalies in health insurance claim data. The evaluation is conducted using metrics such as AUC, accuracy, precision, recall, and F1-score. The results show that the Isolation Forest model performs more consistently and excels in terms of accuracy and F1-score compared to the Autoencoder. Meanwhile, the Autoencoder demonstrates a higher recall in detecting anomalous claims, but its low precision reduces its overall detection effectiveness. Based on these findings, Isolation Forest is recommended as a more reliable method for anomaly detection in this case.

Keywords: Unsupervised Learning, Isolation Forest, Autoencoder, Anomaly Detection, Health Insurance Claims.

1. PENDAHULUAN

Asuransi merupakan produk keuangan yang berfungsi melindungi nasabah dari potensi risiko di masa depan yang tak terduga. Salah satu produk utamanya adalah asuransi kesehatan, yang sangat membantu untuk menutupi biaya pengobatan yang mahal, terutama bagi orang-orang yang tidak memiliki sumber daya finansial yang cukup. Dalam industri asuransi kesehatan, kualitas layanan sangat penting. Seberapa cepat dan

tepat penanganan klaim di rumah sakit menunjukkan layanan yang baik. Penanganan klaim ini harus sesuai dengan janji layanan atau *Service Level Agreement* (SLA) yang dibuat antara klien dan Perusahaan [1].

Penipuan dalam layanan kesehatan merupakan masalah sistemik yang melibatkan berbagai pelaku industri, mulai dari produsen hingga rumah sakit dan dokter. Namun, pihak asuransi, baik pemerintah maupun swasta, menanggung dampak finansial terberat dengan kerugian yang mencapai miliaran setiap tahunnya. Kerugian masif ini bersumber dari berbagai skema klaim palsu, seperti menagih layanan fiktif, melakukan praktik upcoding untuk mendapatkan tagihan lebih tinggi, serta memberikan prosedur medis yang tidak diperlukan secara klinis [2]. Keuntungan finansial sering kali diperoleh secara tidak sah ketika penggugat yang oportunis memanfaatkan peristiwa yang awalnya tidak disengaja untuk mengklaim penyelesaian asuransi, baik melalui jalur hukum maupun di luar pengadilan seperti peristiwa kecelakaan [3].

Metode tradisional untuk mendeteksi penipuan, seperti audit manual dan sistem berbasis aturan, tidak efektif dalam mengatasi kompleksitas penipuan layanan kesehatan kontemporer. Meskipun efektif untuk investigasi yang ditargetkan, audit manual sangat memakan waktu dan banyak pekerjaan, dan tidak cukup untuk mengimbangi jumlah klaim kesehatan yang terus meningkat. Sistem berbasis aturan, di sisi lain, menggunakan ambang batas atau anomali tertentu untuk mengidentifikasi kemungkinan penipuan. Sistem jenis ini, bagaimanapun, seringkali kaku dan tidak fleksibel terhadap pola penipuan yang terus berkembang, dan cenderung menghasilkan banyak alarm palsu atau *False positive* yang dapat membanjiri tim penyelidik dan membuang sumber daya. Meningkatnya jumlah dan variasi klaim layanan kesehatan membuat upaya untuk mengidentifikasi penipuan semakin sulit. Setiap hari, sistem modern menangani jutaan klaim dari berbagai jenis layanan, penyedia, dan kode penagihan [4]. Oleh karena itu penting bagi perusahaan atau pemerintah untuk mendeteksi anomali agar dapat mengambil keputusan yang tepat dan meminimalkan kerugian [5].

Model pembelajaran mesin (ML) telah muncul sebagai solusi yang menguntungkan untuk mendeteksi anomali dalam klaim asuransi. Model ML memiliki kemampuan untuk menganalisis jumlah data yang sangat besar dan menemukan pola kompleks yang tersembunyi, yang biasanya sulit atau tidak mungkin dideteksi oleh manusia [6]. Di antara berbagai pendekatan yang ada, deteksi anomali tanpa pengawasan (*unsupervised anomaly detection*) merupakan kategori yang paling fleksibel dan luas penerapannya. Keunggulan utamanya adalah kemampuannya untuk menganalisis data tanpa memerlukan label apapun yang telah ditentukan sebelumnya [7]. Deteksi anomali memiliki tujuan untuk mengidentifikasi adanya perilaku atau pola data yang berbeda secara signifikan dari kondisi normal sehingga dapat dianggap tidak wajar [8].

Dalam ranah *unsupervised Learning* untuk deteksi anomali, dua model yang menonjol karena pendekatan konseptualnya yang berbeda adalah *Isolation Forest* dan *Autoencoder*. Algoritma *Isolation Forest* dirancang khusus untuk mendeteksi anomali dengan cara memisahkan outlier dari data normal menggunakan struktur pohon keputusan. Metode ini unggul karena mampu mendeteksi anomali secara efisien dan lebih cepat pada dataset besar dibandingkan metode lainnya [9]. Sedangkan kelebihan *Autoencoder* adalah kemampuannya mempelajari fitur-fitur penting dari data secara otomatis, tanpa memerlukan rekayasa fitur yang rumit [10].

Dua penelitian sebelumnya menunjukkan penerapan metode *Isolation Forest* dalam konteks berbeda sebagai solusi deteksi anomali. Penelitian oleh Zulfikar dkk. (2023) menerapkan *Isolation Forest* untuk mendeteksi potensi fraud pada transaksi belanja barang persediaan konsumsi (akun 521811) di satuan kerja Kepolisian Republik Indonesia (BA 060). Hasilnya, model mampu mengidentifikasi anomali transaksi secara

optimal dengan nilai *contamination parameter* 0,3%, sehingga berpotensi dikembangkan sebagai fitur deteksi dini fraud dalam sistem aplikasi keuangan tingkat instansi (SAKTI) [11]. Sementara itu, penelitian Ardiansyah dkk. Pada tahun 2025 mengimplementasikan Isolation Forest pada data jaringan komputer dengan hasil akurasi mencapai 81% pada rasio data latih-uji 80:20. Penelitian ini membuktikan bahwa algoritma tersebut efektif mendeteksi anomali jaringan sekaligus efisien untuk dataset besar, sehingga dapat mendukung peningkatan keamanan jaringan berbasis machine learning [12].

penelitian Sajid Nazir dkk. (2021) menerapkan *Autoencoder* untuk mendeteksi anomali pada sistem SCADA, di mana model yang dilatih dengan data normal mampu mengidentifikasi pola serangan melalui *reconstruction error* sehingga efektif dalam meningkatkan keamanan infrastruktur industri [13]. Selanjutnya, Torabi dkk. (2023) mengusulkan inovasi berbasis *vector reconstruction error* dalam autoencoder untuk jaringan *cloud computing*, yang tidak hanya meningkatkan akurasi deteksi anomali tetapi juga memungkinkan klasifikasi multi-kelas hierarkis dengan kinerja unggul pada metrik akurasi, recall, dan F1-score[14].

Meskipun kedua model ini telah terbukti efektif dalam berbagai domain, penelitian yang secara langsung membandingkan kinerja *Isolation Forest* dan *Autoencoder* pada data klaim asuransi kesehatan. Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis komparatif antara model *Isolation Forest* dan *Autoencoder* untuk deteksi anomali pada dataset klaim asuransi kesehatan. Penelitian ini akan mengevaluasi kinerja kedua model berdasarkan metrik-metrik standar seperti akurasi, presisi, *recall*, dan *F1-score*. Hasil dari penelitian ini diharapkan dapat memberikan wawasan berharga bagi perusahaan asuransi dalam memilih dan mengimplementasikan model *unsupervised Learning* yang paling efektif dan efisien untuk melindungi aset mereka dan menjaga integritas sistem layanan kesehatan.

2. TINJAUAN PUSTAKA

2.1 Deteksi Anomali

Deteksi anomali adalah sebuah teknik untuk mengidentifikasi observasi atau item data yang polanya menyimpang secara signifikan dari perilaku yang dianggap normal. Penyimpangan ini sering kali menjadi indikator adanya aktivitas tidak wajar atau abnormal, seperti yang terjadi dalam kasus insiden keamanan siber maupun upaya penipuan pada transaksi kartu kredit [15].

2.2 Asuransi Kesehatan

Asuransi kesehatan merupakan sebuah instrumen proteksi keuangan yang menyediakan pertanggungan untuk biaya pengobatan dan perawatan medis ketika pemegangnya jatuh sakit atau mengalami kecelakaan. Manfaatnya adalah membantu meringankan tanggungan biaya medis yang besar dan tidak terduga, serta memfasilitasi akses ke fasilitas kesehatan yang lebih berkualitas, khususnya bagi mereka yang tidak sanggup menanggungnya sendiri [16].

2.3 Isolation Forest

Isolation Forest adalah metode *unsupervised machine Learning* untuk deteksi anomali yang menggunakan sekumpulan pohon keputusan (*decision trees*). Prinsip kerjanya didasarkan pada asumsi bahwa data anomali lebih mudah untuk diisolasi atau dipisahkan daripada data normal. Proses ini terdiri dari dua tahap utama: pelatihan untuk membangun pohon isolasi, dan evaluasi untuk menghitung skor anomali pada setiap data berdasarkan seberapa cepat ia berhasil dipisahkan. Berkat pendekatan ini, keunggulan utama *Isolation Forest* adalah efisiensinya yang tinggi dan kecepatannya dalam

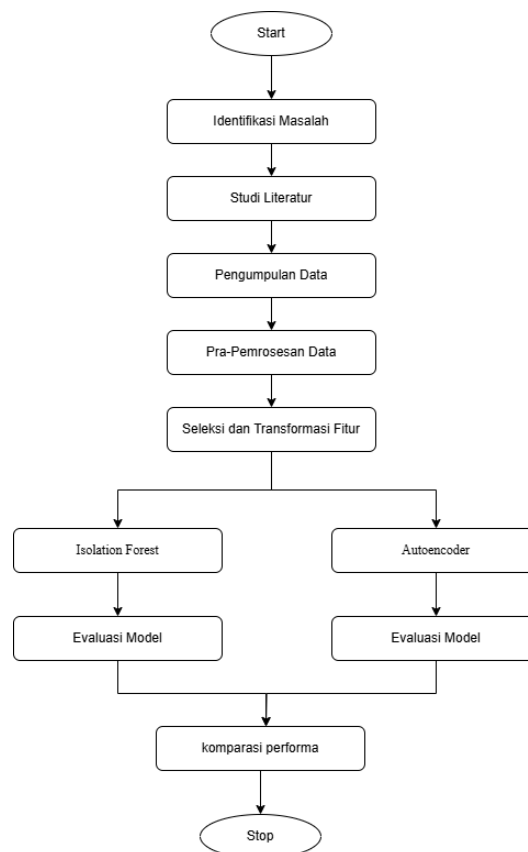
mengidentifikasi outlier, karena anomali memerlukan lebih sedikit langkah untuk diisolasi [17].

2.4 Autoencoder

Autoencoder adalah model jaringan saraf yang dirancang untuk mereplikasi inputnya pada bagian output melalui proses kompresi (oleh *Encoder*) dan dekomposisi (oleh *Decoder*). Dengan arsitektur berbentuk jam pasir, model ini secara efektif mereduksi dimensi data pada lapisan tengahnya (*code layer*) untuk menangkap fitur-fitur non-linear dan pola tersembunyi. Keunggulan utama metode ini terletak pada kemampuannya untuk belajar fitur secara hierarkis, sehingga mampu menghasilkan representasi data yang optimal dan semakin abstrak di tiap tingkatan [18].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental komparatif dengan menerapkan metode *unsupervised machine Learning*. Fokus utama dari penelitian ini adalah membandingkan performa dua algoritma deteksi anomali, yaitu *Isolation Forest* dan *Autoencoder*, dalam mengidentifikasi klaim asuransi kesehatan yang tergolong abnormal atau berpotensi *fraud*. Penelitian ini disusun melalui tahapan-tahapan sebagai berikut.



Gambar 1 Kerangka Penelitian

1. Identifikasi Masalah

Peneliti merumuskan permasalahan utama, yaitu bagaimana mendeteksi anomali atau potensi *fraud* pada klaim asuransi kesehatan secara otomatis menggunakan

pendekatan *unsupervised Learning*. Masalah ini menjadi penting karena banyaknya kasus penyalahgunaan klaim yang sulit dideteksi secara manual.

2. Studi Literatur

Kajian literatur dilakukan untuk memperoleh landasan teori yang kuat mengenai deteksi anomali, karakteristik data klaim asuransi, serta algoritma *unsupervised Learning*, khususnya *Isolation Forest* dan *Autoencoder*. Kajian ini juga bertujuan menemukan celah penelitian (*research gap*) yang belum banyak dijelajahi oleh penelitian sebelumnya.

3. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diambil dari situs Kaggle dengan judul Healthcare Provider Fraud Detection Analysis, yang tersedia di tautan: <https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis/data>. Dataset ini berisi informasi klaim kesehatan dari berbagai penyedia layanan, termasuk variable.

4. Pra-Pemrosesan Data

Data yang telah dikumpulkan selanjutnya melalui proses pembersihan dari nilai kosong (*missing values*), duplikasi data, dan nilai ekstrem (*outlier*). Kemudian dilakukan transformasi data, seperti normalisasi terhadap fitur numerik dan one-hot encoding terhadap fitur kategorikal agar data dapat diproses secara optimal oleh model pembelajaran mesin[19].

5. Seleksi dan Transformasi Fitur

Fitur yang tidak relevan dieliminasi, dan fitur utama yang berkaitan dengan pola klaim tetap dipertahankan. Jika diperlukan, dilakukan teknik reduksi dimensi seperti *Principal Component Analysis* (PCA) untuk menyederhanakan struktur data tanpa menghilangkan informasi penting[20].

6. Penerapan Model: *Isolation Forest* dan *Autoencoder*

Data yang telah diproses kemudian digunakan untuk membangun dua model deteksi anomaly yaitu *Isolation Forest* dan *Autoencoder*. *Isolation Forest*, yaitu model berbasis pohon acak yang efektif dalam mengisolasi data anomali lebih cepat dibanding data normal. *Autoencoder*, yaitu jaringan saraf tiruan dengan struktur encoder-decoder yang mendeteksi anomali berdasarkan selisih (*error*) antara input dan output hasil rekonstruksi.

7. Evaluasi Model (Masing-masing)

Masing-masing model dievaluasi menggunakan metrik seperti reconstruction error untuk *Autoencoder* dan anomaly score untuk *Isolation Forest*. Jika tersedia label *fraud* (*PotentialFraud*), evaluasi tambahan dilakukan menggunakan metrik *precision*, *recall*, dan ROC-AUC untuk mengukur seberapa baik model dalam mengenali klaim abnormal.

8. Komparasi Performa

Hasil dari kedua model dibandingkan dari berbagai aspek, seperti akurasi relatif dalam mendeteksi *anomaly*. Tujuan tahap ini adalah menentukan model yang paling efektif dan efisien untuk diterapkan dalam konteks klaim asuransi kesehatan.

4. HASIL DAN PEMBAHASAN

4.1 Pengolahan Dataset

Dataset yang digunakan dalam penelitian ini diambil dari situs Kaggle dengan judul Healthcare Provider Fraud Detection Analysis, yang tersedia di tautan: <https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis/data>. Data yang sudah dikumpulkan di proses pada Tahap pra-pemrosesan dan agregasi data bertujuan untuk mengubah data klaim mentah menjadi representasi yang lebih bermakna dalam bentuk profil penyedia layanan (*provider*). Data klaim yang berasal

dari dua sumber, yaitu rawat inap (*inpatient*) dan rawat jalan (*outpatient*), pertama-tama digabungkan menjadi satu dataset komprehensif. Langkah ini penting karena memungkinkan analisis dilakukan secara menyeluruh terhadap seluruh aktivitas klaim yang dilakukan oleh masing-masing provider.

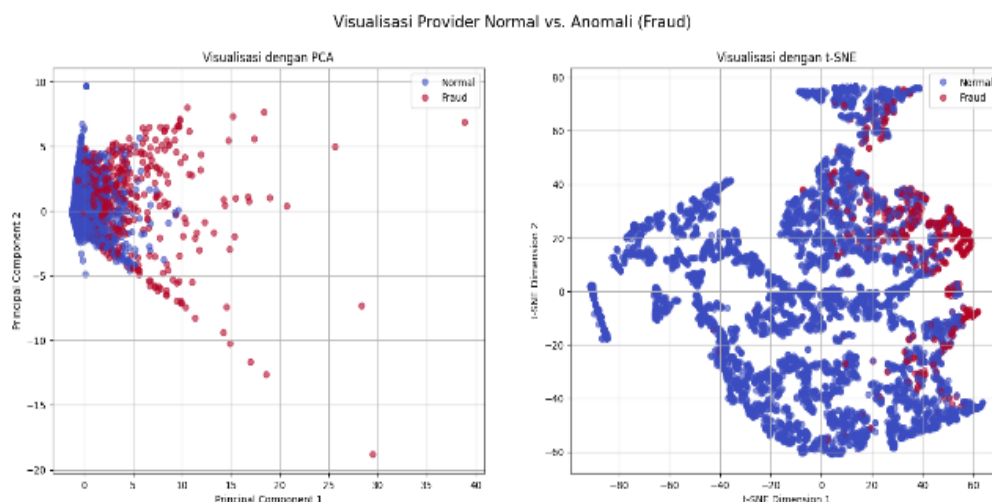
Setelah penggabungan data, dilakukan proses agregasi dengan menghitung berbagai indikator operasional per provider. Selanjutnya, dari hasil agregasi tersebut dibentuk sejumlah fitur turunan (*engineered features*) yang bertujuan untuk mempermudah identifikasi pola anomali. Akhir dari tahap ini ditandai dengan penggabungan hasil agregasi profil provider dengan informasi label 'PotentialFraud' yang tersedia dalam dataset. Meskipun model yang digunakan dalam penelitian ini merupakan model *unsupervised* (tidak memanfaatkan label selama pelatihan), informasi tersebut tetap berguna untuk evaluasi performa model pada tahap selanjutnya. Dengan demikian, tahap pra-pemrosesan dan rekayasa fitur ini membentuk fondasi yang kuat untuk penerapan model deteksi anomali secara efisien dan akurat.

4.2 Seleksi Fitur

Pada tahap ini, fitur Provider dan label PotentialFraud dihapus dari data, sehingga hanya fitur numerik yang digunakan sebagai input model. Label PotentialFraud dikonversi menjadi nilai biner, di mana 'Yes' menjadi 1 (*fraud*) dan 'No' menjadi 0 (normal). Selanjutnya, seluruh fitur diskalakan menggunakan *StandardScaler* agar memiliki rata-rata 0 dan standar deviasi 1, sehingga model dapat mendeteksi pola anomali secara lebih optimal tanpa bias skala antar fitur.

4.3 Visualisasi Anomali

Visualisasi ini menggunakan dua metode reduksi dimensi, PCA dan t-SNE, untuk memproyeksikan data berdimensi tinggi ke dalam dua dimensi guna mengamati pola antara provider normal dan yang terindikasi *fraud*. Hasil dari PCA menunjukkan distribusi berdasarkan dua komponen utama, sementara t-SNE, yang lebih sensitif terhadap struktur lokal, mampu memperlihatkan pemisahan yang lebih jelas antara provider *fraud* dan non-*fraud*. Titik-titik berwarna merah (*fraud*) cenderung membentuk kelompok tersendiri, menandakan adanya perbedaan karakteristik perilaku klaim dibandingkan dengan titik biru (normal). Visualisasi ini memperkuat dugaan bahwa data *fraud* memiliki pola yang dapat dikenali secara visual, mendukung penerapan model deteksi anomali.



Gambar 2 Visualisasi Anomali

4.4 Hasil Isolation Forest

Hasil evaluasi model *Isolation Forest* menunjukkan bahwa model ini cukup efektif dalam mendeteksi anomali pada klaim asuransi kesehatan. Nilai AUC (Area Under Curve) sebesar 0.7667 mengindikasikan bahwa model memiliki kemampuan yang baik dalam membedakan antara klaim normal dan *fraud*. Sementara itu, nilai *F1-score* untuk kelas *fraud* sebesar 0.5771 menunjukkan keseimbangan moderat antara *precision* dan *recall* dalam mendeteksi klaim yang terindikasi sebagai penipuan.

Dari laporan klasifikasi, model mampu mendeteksi klaim normal dengan sangat baik, ditunjukkan oleh *precision* dan *recall* sebesar 0.96, yang menghasilkan *F1-score* tinggi (0.96) untuk kelas normal. Namun, untuk kelas *fraud*, model hanya memiliki *precision* 0.58 dan *recall* 0.58, yang berarti hanya sekitar 58% klaim *fraud* yang berhasil dikenali, dan dari semua klaim yang diprediksi sebagai *fraud*, 58% benar-benar *fraud*. Akurasi keseluruhan mencapai 92%, namun ini dipengaruhi oleh ketimpangan kelas, karena mayoritas data termasuk dalam kelas normal. Secara keseluruhan, *Isolation Forest* menunjukkan kinerja yang cukup baik dalam konteks *unsupervised Learning*, terutama dalam menjaga akurasi dan tetap mengenali sebagian besar klaim *fraud*.

4.5 Hasil Autoencoder

Model *Autoencoder* menunjukkan performa yang cukup baik dalam mendeteksi anomali dengan AUC score sebesar 0.7790, menandakan kemampuan model yang baik dalam membedakan antara data normal dan *fraud*. Nilai *F1-score* keseluruhan sebesar 0.5008 mencerminkan keseimbangan sedang antara presisi dan *recall*, terutama dalam kasus *fraud* yang sulit dikenali karena jumlahnya jauh lebih sedikit. Akurasi keseluruhan mencapai 88%, yang cukup tinggi, tetapi kurang mencerminkan kinerja dalam mendeteksi *fraud* karena data tidak seimbang.

Dari laporan klasifikasi, model memiliki *precision* tinggi (0.96) dan *recall* (0.90) pada kelas normal, artinya model sangat efektif mengenali klaim asuransi yang sah. Namun, untuk kelas *fraud*, *precision* hanya 0.40 dan *recall* 0.66, menunjukkan model cukup berhasil menemukan *fraud* tetapi juga menghasilkan cukup banyak *False positive*. Ini menunjukkan bahwa meskipun model sudah mampu menangkap sebagian besar anomali, masih ada ruang untuk perbaikan terutama dalam mengurangi kesalahan prediksi pada kelas minoritas.

4.6 Tabel Perbandingan Model

Berikut ini adalah tabel perbandingan antara model *Isolation Forest* dan *Autoencoder* berdasarkan berbagai metrik evaluasi yang digunakan untuk menilai kinerja dalam mendeteksi anomaly provider klaim asuransi kesehatan

Tabel 1 Perbandingan Model

Metrik Evaluasi	<i>Isolation Forest</i>	<i>Autoencoder</i>
AUC Score	0,7667	0,779
Accuracy	0,92	0,88
<i>F1-score</i> (Total)	0,5771	0,5008
<i>Precision</i> - Normal	0,96	0,96
<i>Recall</i> - Normal	0,96	0,90
<i>F1-score</i> - <i>Fraud</i>	0,58	0,50
<i>Precision</i> - <i>Fraud</i>	0,58	0,40
<i>Recall</i> - <i>Fraud</i>	0,58	0,66
Macro Avg F1	0,77	0,72

Metrik Evaluasi	<i>Isolation Forest</i>	<i>Autoencoder</i>
Weighted Avg F1	0,92	0,89

4.7 Komparasi Performa

Berdasarkan hasil evaluasi model, dapat disimpulkan bahwa *Isolation Forest* merupakan model yang lebih unggul dibandingkan *Autoencoder* dalam mendeteksi *fraud* pada studi kasus ini. Meskipun *Autoencoder* memiliki nilai AUC yang sedikit lebih tinggi (0.7790 dibandingkan 0.7667) dan *recall* yang lebih baik dalam mendeteksi *fraud* (0.66 vs 0.58), namun *Isolation Forest* menunjukkan performa yang lebih seimbang secara keseluruhan. Hal ini terlihat dari nilai *F1-score* yang lebih tinggi (0.5771 vs 0.5008), *precision* yang lebih baik pada kelas *fraud* (0.58 vs 0.40), serta akurasi total yang lebih tinggi (0.92 vs 0.88). Tingginya *precision* pada *Isolation Forest* sangat penting untuk mengurangi *False positive*, yang dapat berdampak negatif dalam sistem deteksi *fraud* nyata. Oleh karena itu, *Isolation Forest* direkomendasikan sebagai model yang lebih andal dan efisien untuk mendeteksi *fraud* dalam konteks ini, karena memberikan keseimbangan optimal antara akurasi, presisi, dan stabilitas performa di kedua kelas.

5. KESIMPULAN

Isolation Forest lebih unggul secara keseluruhan, dengan akurasi, *F1-score*, dan *precision* yang lebih tinggi, menjadikannya lebih andal untuk mendeteksi klaim anomali secara tepat dan konsisten. Sedangkan *Autoencoder* memiliki kelebihan dalam *recall* pada klaim *fraud*, namun *precision* yang rendah membuatnya kurang efektif untuk penggunaan yang memerlukan akurasi tinggi dalam mendeteksi penipuan.

UCAPAN TERIMAKASIH

Terima kasih penulis sampaikan kepada pihak institusi yang telah memberikan fasilitas dan dukungan, para dosen pembimbing serta rekan peneliti yang telah membantu melalui arahan dan diskusi konstruktif, serta keluarga dan sahabat yang senantiasa memberikan doa dan dukungan moril sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] Q. yumansyah Qori, A. Turmudi Zy, and M. Fatchan, "Prediksi Jumlah Kasus Klaim Indemnity Dengan Menggunakan Algoritma Regresi Linear Pada Asuransi Mandiri Inhealth," *Bull. Inf. Technol.*, vol. 4, no. 3, pp. 299–305, 2023, doi: 10.47065/bit.v4i3.733.
- [2] T. K. Mackey, K. Miyachi, D. Fung, S. Qian, and J. Short, "Combating health care fraud and abuse: Conceptualization and prototyping study of a blockchain antifraud framework," *J. Med. Internet Res.*, vol. 22, no. 9, 2020, doi: 10.2196/18623.
- [3] N. Kumaraswamy, M. K. Markey, T. Ekin, J. C. Barner, and K. Rascati, "Healthcare Fraud Data Mining Methods: A Look Back and Look Ahead," *Perspect. Heal. Inf. Manag.*, vol. 19, no. 1, 2022.
- [4] T. Azad and P. William, "Fraud detection in healthcare billing and claims Fraud detection in healthcare billing and claims," *Int. J. Sci. Res. Arch.*, vol. 13, no. 2, 2024, doi: 10.30574/ijrsra.2024.13.2.2606.
- [5] D. E. Prasetyo, G. A. A. Wulandari, Z. Meini, and Fauziah, "Identifikasi Fraud Dalam Pemeriksaan Internal Melalui Data Analytics," *J. EQUITY*, vol. 26, no. 1, pp. 78–98, 2024, doi: 10.34209/equ.v26i1.7418.
- [6] P. K. Baruah *et al.*, "Integrating Machine Learning Models with Business Rule Triggers to Boost Performance in Health Insurance Fraud Detection : A Case Study,"

- Var. Adv. Sci. Risk*, vol. 18, 2025.
- [7] W. Hilal, S. A. Gadsden, and J. Yawney, "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances," *Expert Syst. Appl.*, vol. 193, p. 116429, 2022, doi: 10.1016/j.eswa.2021.116429.
- [8] G. S. W. Prabuningrat, D. P. Hostiadi, and N. L. P. Srinadi, "Klasifikasi Deteksi Anomali Menggunakan Metode Machine Learning," *Semin. Has. Penelit. Inform. dan Komput.* 2024, vol. 1, no. 2, p. 2024, 2024, [Online]. Available: <http://www.jubikom.unpak.ac.id>
- [9] D. B. Santoso and Y. Wahyuni, "Sistem Log Web Server Sebagai Pendeteksi Anomali Menggunakan Isolation Forest," *J. Apl. Bisnis dan Komput.*, vol. 4, no. 3, pp. 2807–5986, 2024, [Online]. Available: <http://www.jubikom.unpak.ac.id>
- [10] M. F. Dzulqarnain, A. Fadlil, and I. Riadi, "Implementasi CNN berbasis Autoencoder pada Klasifikasi Pola Batik," *Proceeding Informatics Collab. Dessimation Meet. Infocoding 2025*, pp. 189–194, 2025.
- [11] A. Zulfikar, F. A. Rahmani, and N. Azizah, "Deteksi Anomali Menggunakan Isolation Forest Belanja Barang Persediaan Konsumsi Pada Satuan Kerja Kepolisian Republik Indonesia," *J. Manaj. Perbendaharaan*, vol. 4, no. 1, pp. 1–15, 2023, doi: 10.33105/jmp.v4i1.435.
- [12] R. Ardiansyah, L. Sunardi, P. S. Informatika, U. B. Insan, K. Jaringan, and I. Forest, "Implementasi Metode Isolation Forest Untuk Deteksi Anomali Dalam Data Jaringan," *Semin. Nas. ESCAF (Economic, Soc. Sci. Comput. Agric. Fish. 4th)*, pp. 933–939, 2025.
- [13] S. Nazir, S. Patel, and D. Patel, "Autoencoder Based Anomaly Detection for SCADA Networks," *Int. J. Artif. Intell. Mach. Learn.*, vol. 11, no. 2, pp. 83–99, 2021, doi: 10.4018/ijaiml.20210701.0a6.
- [14] H. Torabi, S. L. Mirtaheri, and S. Greco, "Practical autoencoder based anomaly detection by using vector reconstruction error," *Cybersecurity*, vol. 6, no. 1, pp. 1–13, 2023, doi: 10.1186/s42400-022-00134-9.
- [15] D. R. K. Saputra, Y. V. Via, and A. N. Sihananto, "Deteksi Anomali Menggunakan Ensemble Learning Dan Random Oversampling Pada Penipuan Transaksi Keuangan," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2779–2788, 2024, doi: 10.23960/jitet.v12i3.4910.
- [16] A. Jayadie *et al.*, *Pembiayaan Kesehatan*, no. September. Bandung: MEDIA SAINS INDONESIA, 2023.
- [17] R. P. Dewa and Windarto, "Deteksi Anomali Jaringan Menggunakan Isolation Forest pada Log Wazuh dengan Pemberitahuan WhatsApp di PT XYZ Network Anomaly Detection Using Isolation Forest on Wazuh Logs with WhatsApp Notifications at PT XYZ," *KRESNA J. Ris. dan Pengabd. Masy.*, vol. 4, pp. 208–216, 2024.
- [18] F. S. Nugraha and H. F. Pardede, "Autoencoder untuk Sistem Prediksi Berat Lahir Bayi," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 235–244, 2022, doi: 10.25126/jtiik.2022923868.
- [19] I. Rianto and I. P. I. Santosa, *Data Preparation untuk Machine Learning & Deep Learning*. Penerbit Andi, 2025.
- [20] W. R. Nurputra, W. Witanti, and A. Komarudin, "Principal Component Analysis (PCA) Untuk Meningkatkan Hasil Klasterisasi Penjualan Video Game Menggunakan Algoritma K-Means," *J. LOCUS Penelit. Pengabd.*, vol. 4, no. 8, pp. 4683–4703, 2025.

