Pengembangan Model Artificial Intelligence untuk Mendeteksi Anomali dalam Rekam Medis Pasien sebagai Upaya Pencegahan Kesalahan Klinis

Rahmadyo Yudhi Prabowo[1*], Gunawan[2], Sandri Kesuma[3]
Rekam Medis dan Informasi Kesehatan, Politeknik Kesehatan Kemenkes Malang[1,2,3]
*e-mail: rahmadyo_yudhi@poltekkes-malang.ac.id

ABSTRAK

Kesalahan klinis yang bersumber dari ketidakwajaran data rekam medis masih menjadi tantangan signifikan dalam sistem pelayanan kesehatan, karena berpotensi memengaruhi akurasi pengambilan keputusan medis dan keselamatan pasien. Ketidakwajaran tersebut dapat berupa kesalahan pencatatan nilai klinis, ketidaksesuaian dosis obat, maupun inkonsistensi informasi pasien. Penelitian ini bertujuan untuk mengembangkan dan menguji model Artificial Intelligence (AI) berbasis *unsupervised anomaly detection* dalam mendeteksi anomali pada rekam medis pasien sebagai mekanisme pencegahan dini terhadap kesalahan klinis. Metode Isolation Forest digunakan untuk mempelajari pola mayoritas data dan mengidentifikasi observasi yang menyimpang tanpa memerlukan pelabelan kesalahan sebelumnya. Dataset penelitian berasal dari data rekam medis pasien rawat jalan yang mencakup variabel demografis, diagnosis, nilai laboratorium, dosis obat, serta kelengkapan resume medis. Hasil analisis menunjukkan bahwa model mampu mengidentifikasi sekitar 10% data sebagai anomali dengan skor penyimpangan yang relatif tinggi. Visualisasi hasil memperlihatkan pemisahan yang jelas antara data normal dan data anomali, sekaligus mengindikasikan adanya kombinasi nilai klinis yang tidak lazim dan berpotensi menimbulkan risiko klinis. Temuan ini menegaskan bahwa model AI berbasis Isolation Forest memiliki potensi kuat untuk dikembangkan sebagai sistem peringatan dini guna mendukung keselamatan pasien dan meningkatkan kualitas pengelolaan rekam medis dalam layanan kesehatan.

Kata kunci : *Artificial Intelligence*; Deteksi Anomali; *Isolation Forest*; Kesalahan Klinis; Rekam Medis

ABSTRACT

Clinical errors arising from irregularities in medical record data remain a major challenge in healthcare systems, as they may compromise clinical decision-making accuracy and patient safety. Such irregularities may include incorrect recording of clinical values, inappropriate medication dosages, and inconsistencies in patient information. This study aims to develop and evaluate an Artificial Intelligence (AI) model based on unsupervised anomaly detection to identify anomalies in patient medical records as an early prevention mechanism for clinical errors. The Isolation Forest method was employed to learn the dominant data patterns and detect anomalous observations without requiring prior error labeling. The dataset was derived from outpatient medical records, encompassing demographic variables, diagnoses, laboratory values, medication dosages, and completeness of medical summaries. The results indicate that the model successfully identified approximately 10% of the data as anomalies with relatively high anomaly scores. Visualization results demonstrate a clear separation between normal and anomalous data, revealing unusual combinations of clinical values that may pose potential clinical risks. These findings suggest that an Isolation Forest–based AI model holds substantial potential as an early warning system to enhance patient safety and improve the quality of medical record management in healthcare services.

Keywords **:** *Artificial Intelligence; Anomaly Detection; Clinical Errors; Isolation Forest; Medical Records*

PENDAHULUAN

Rekam medis merupakan elemen fundamental dalam sistem pelayanan kesehatan modern karena berfungsi sebagai landasan utama dalam proses diagnosis, penentuan terapi, pemantauan perjalanan penyakit, serta evaluasi luaran klinis pasien. Keakuratan, kelengkapan, dan konsistensi data rekam medis berimplikasi langsung terhadap keselamatan pasien dan mutu pengambilan keputusan klinis. Organisasi

Kesehatan Dunia menegaskan bahwa kualitas informasi klinis merupakan faktor kunci dalam upaya pencegahan insiden keselamatan pasien (World Health Organization [WHO], 2017). Namun demikian, dalam praktik pelayanan kesehatan sehari-hari, rekam medis masih rentan mengandung kesalahan teknis dan ketidakwajaran data, baik berupa kesalahan pencatatan nilai klinis, ketidaksesuaian dosis obat, maupun inkonsistensi data administratif, yang dipengaruhi oleh kompleksitas informasi klinis, keterbatasan waktu, serta faktor manusia dalam proses dokumentasi (Carayon et al., 2014).

Kesalahan klinis tidak selalu muncul sebagai akibat dari pertimbangan medis yang keliru, melainkan sering kali dipicu oleh kualitas data yang menjadi dasar pengambilan keputusan tersebut. Bates et al. (1999) menunjukkan bahwa kesalahan informasi, khususnya pada sistem pemesanan dan dokumentasi medis, berkontribusi signifikan terhadap terjadinya medication error dan risiko klinis lainnya. Pemeriksaan manual terhadap data rekam medis membutuhkan ketelitian tinggi, konsistensi, dan sumber daya manusia yang memadai. Dalam lingkungan pelayanan kesehatan dengan beban kerja tinggi, pendekatan manual menjadi tidak efisien dan berisiko menyebabkan anomali data terlewatkan, sehingga memunculkan kebutuhan akan sistem pendukung berbasis teknologi yang mampu melakukan deteksi ketidakwajaran data secara sistematis, objektif, dan berkelanjutan.

Perkembangan Artificial Intelligence (AI), khususnya pada bidang machine learning, membuka peluang besar untuk memanfaatkan data rekam medis sebagai sumber pembelajaran pola klinis dalam skala besar. Beam dan Kohane (2018) serta Topol (2019) menegaskan bahwa pemanfaatan big data dan machine learning dalam kesehatan berpotensi meningkatkan akurasi klinis dan keselamatan pasien. Salah satu pendekatan yang relevan dalam konteks ini adalah deteksi anomali (*anomaly detection*), yang bertujuan mengidentifikasi data yang menyimpang dari pola mayoritas. Pendekatan ini sangat sesuai untuk data rekam medis karena kesalahan klinis relatif jarang terjadi dan umumnya tidak memiliki label eksplisit, sehingga pendekatan supervised learning menjadi kurang optimal (Chandola et al., 2009; Ruff et al., 2021).

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan model Artificial Intelligence berbasis *unsupervised anomaly detection* menggunakan algoritma Isolation Forest yang diperkenalkan oleh Liu et al. (2008). Model ini dirancang untuk mendeteksi anomali dalam data rekam medis pasien sebagai bentuk pencegahan dini terhadap kesalahan klinis dan kesalahan teknis pencatatan data. Berbeda dengan pendekatan berbasis aturan yang bersifat statis, Isolation Forest mampu menangkap pola multivariat yang kompleks pada data berdimensi tinggi. Penelitian ini tidak hanya berfokus pada aspek teknis deteksi anomali, tetapi juga menekankan interpretasi klinis dan visualisasi hasil agar mudah dipahami oleh tenaga kesehatan, serta membuka peluang integrasi model sebagai modul analitik dalam Sistem Informasi Manajemen Rumah Sakit (SIMRS) untuk mendukung validasi data dan sistem peringatan dini secara berkelanjutan (Sendak et al., 2020; Rajpurkar et al., 2022).

METODE PENELITIAN

Dataset dan Variabel

Dataset penelitian terdiri dari 200 data rekam medis pasien rawat jalan yang telah dianonimkan. Variabel yang digunakan meliputi usia pasien, jenis kelamin, kode diagnosis (ICD-10), nilai hemoglobin, dosis obat, dan kelengkapan resume medis. Rentang usia pasien berada pada 15 hingga 84 tahun dengan rata-rata usia 50,9 tahun. Nilai hemoglobin dan dosis obat menunjukkan variasi yang cukup lebar, sehingga relevan untuk penerapan pendekatan deteksi anomali berbasis unsupervised learning. Beberapa contoh data dapat dilihat sebagai berikut :

Tabel 1. Data Anomali yang Terdeteksi

No	Usia(th)	Jenis Kelamin	Hb	Dosis Obat	Diagnosis	Resume	Skor	Kategori

			(g/dl)	(mg)	(ICD-10)	Medis	Anomali	
1	45	L	13,5	50	I10	Lengkap	0,12	Normal
2	52	P	11,8	75	E11	Lengkap	0,18	Normal
3	60	L	10,2	150	J18	Lengkap	0,31	Anomali Ringan
4	38	P	9,5	500	D50	Tidak lengkap	0,47	Anomali sedang
5	70	L	8,9	1200	I10	Lengkap	0,63	Anomali sedang

Pra-pemrosesan Data

Variabel numerik dinormalisasi menggunakan StandardScaler, sedangkan variabel kategorikal dikonversi menjadi representasi numerik menggunakan One-Hot Encoding. Tahap pra-pemrosesan ini bertujuan untuk memastikan seluruh variabel dapat diproses secara optimal oleh model.

Model *Isolation Forest*

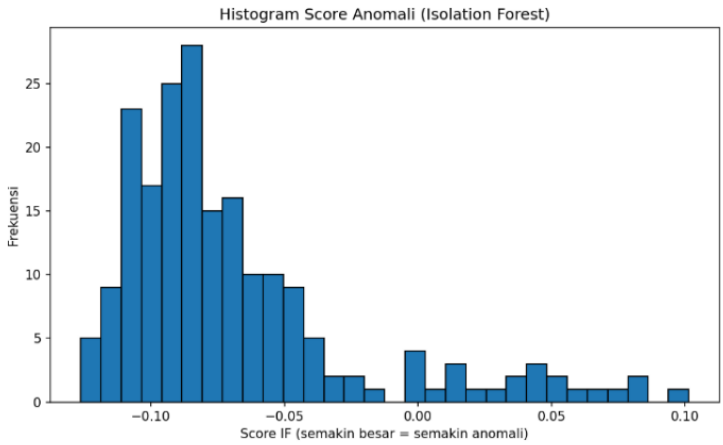
Model *Isolation Forest* dibangun dengan sejumlah pohon isolasi dan parameter contamination sebesar 10%, yang merepresentasikan asumsi proporsi data anomali. Model dilatih secara unsupervised untuk mempelajari pola mayoritas data tanpa memerlukan label kesalahan.

Implementasi Menggunakan Python

Penelitian ini diimplementasikan menggunakan bahasa pemrograman Python, yang dipilih karena ekosistem pustaka machine learning yang matang dan banyak digunakan dalam pengembangan sistem analitik kesehatan. Python digunakan untuk seluruh tahapan penelitian, mulai dari pra-pemrosesan data, pelatihan model, hingga visualisasi hasil. Pra-pemrosesan data dilakukan menggunakan pustaka Pandas untuk pengelolaan data tabular dan Scikit-learn untuk normalisasi data serta pengkodean variabel kategorikal. Model Isolation Forest diimplementasikan menggunakan modul sklearn.ensemble, yang menyediakan algoritma unsupervised anomaly detection yang efisien dan dapat diskalakan. Visualisasi hasil penelitian, seperti histogram skor anomali, proporsi kasus normal dan anomali, serta sebaran hubungan nilai hemoglobin dan dosis obat, dihasilkan menggunakan pustaka Matplotlib. Visualisasi ini berperan penting dalam mendukung interpretasi hasil oleh tenaga kesehatan dan pemangku kepentingan non-teknis.

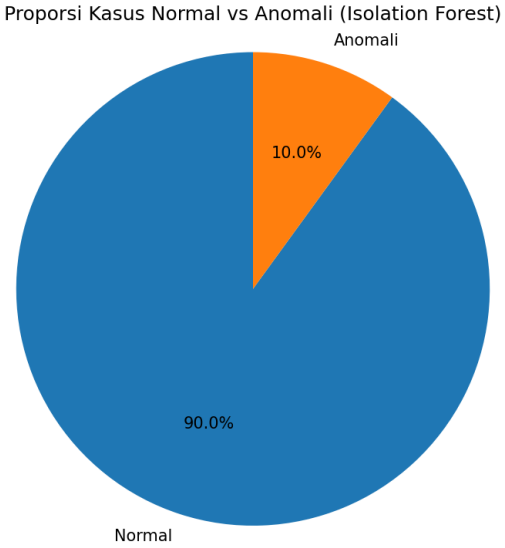
HASIL DAN PEMBAHASAN

Bagian ini membahas hasil penerapan model Artificial Intelligence berbasis Isolation Forest dalam mendeteksi anomali pada data rekam medis pasien. Pembahasan difokuskan pada interpretasi hasil visualisasi, relevansinya terhadap pencegahan kesalahan klinis, serta potensi penerapannya dalam sistem informasi rumah sakit. Pendekatan ini bertujuan menjembatani hasil teknis model AI dengan konteks praktis pengelolaan rekam medis dan keselamatan pasien. Hasil penerapan model anomaly ini dapat divisualisasikan dalam bentuk grafik sebagai berikut :



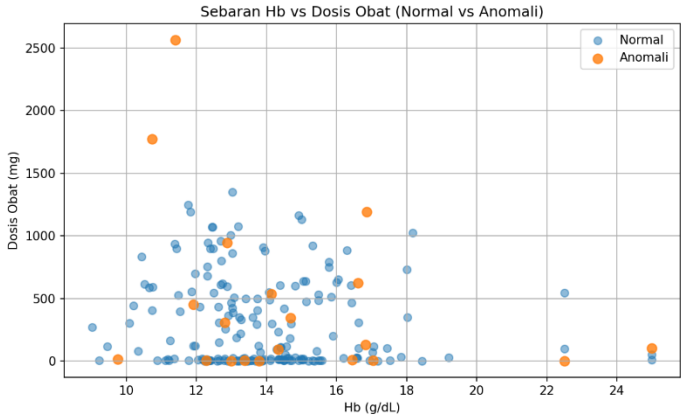
Gambar 1. Distribusi Skor Anomali

Distribusi skor anomali yang ditunjukkan pada gambar 1 menunjukkan bahwa sebagian besar data rekam medis berada pada kelompok normal dengan skor anomali rendah, sementara sebagian kecil data memiliki skor anomali tinggi yang mengindikasikan penyimpangan signifikan terhadap pola mayoritas. Pola ini menunjukkan bahwa model Isolation Forest mampu mempelajari karakteristik data normal secara efektif dan mengisolasi data yang menyimpang. Dari perspektif klinis, data dengan skor anomali tinggi berpotensi merepresentasikan ketidakwajaran nilai klinis atau kesalahan teknis pencatatan yang perlu ditinjau lebih lanjut.



Gambar 2. Proporsi Data Normal

Proporsi data normal dan anomali yang ditunjukkan pada gambar 2 menunjukkan bahwa sekitar 90% data terklasifikasi sebagai normal dan 10% sebagai anomali. Proporsi ini sesuai dengan karakteristik data klinis, di mana kesalahan atau ketidakwajaran relatif jarang terjadi. Penetapan proporsi anomali yang terbatas penting untuk mencegah terjadinya alarm fatigue, sehingga sistem peringatan dini yang dikembangkan tetap selektif dan relevan bagi tenaga kesehatan.



Gambar 3. Grafik Sebaran Data

Analisis sebaran hubungan antara nilai hemoglobin dan dosis obat menunjukkan bahwa data yang diklasifikasikan sebagai normal cenderung terkonsentrasi pada rentang nilai yang lazim secara klinis, sedangkan data anomali tersebar pada kombinasi nilai yang tidak umum, baik berupa nilai ekstrem maupun ketidaksesuaian antarvariabel. Pola ini mengindikasikan bahwa ketidakwajaran data rekam medis tidak selalu muncul sebagai penyimpangan tunggal pada satu variabel, melainkan sebagai hasil interaksi multivariat yang kompleks. Temuan ini sejalan dengan pandangan Chandola et al. (2009) dan Ruff et al. (2021) yang menegaskan bahwa anomali dalam data berdimensi tinggi sering kali tidak dapat diidentifikasi secara akurat melalui pendekatan berbasis ambang batas (rule-based), karena pendekatan tersebut bersifat statis dan mengabaikan hubungan antarvariabel. Dengan demikian, visualisasi hubungan hemoglobin dan dosis obat dalam penelitian ini memperkuat argumen bahwa pendekatan unsupervised anomaly detection lebih sesuai untuk konteks data klinis yang heterogen dan dinamis.

Keunggulan model Isolation Forest dalam penelitian ini terletak pada kemampuannya mengisolasi data yang menyimpang berdasarkan struktur distribusi data secara keseluruhan, bukan semata-mata pada nilai ekstrem individual. Pendekatan ini konsisten dengan konsep yang diperkenalkan oleh Liu et al. (2008), bahwa data anomali cenderung lebih mudah “diisolasi” dibandingkan data normal karena memiliki karakteristik yang jarang muncul. Dalam konteks rekam medis, hal ini menjadi sangat relevan karena kombinasi nilai laboratorium dan dosis obat yang tidak lazim dapat merepresentasikan potensi kesalahan pencatatan, ketidaksesuaian terapi, atau kondisi klinis yang memerlukan perhatian khusus. Dibandingkan penelitian terdahulu yang menerapkan Isolation Forest pada data laboratorium atau klaim asuransi secara terpisah (Beam & Kohane, 2018; Krittanawong et al., 2017), penelitian ini menempatkan hasil deteksi anomali secara eksplisit dalam kerangka pencegahan kesalahan klinis, sehingga memperluas relevansi klinis dari pendekatan teknis yang digunakan.

Lebih lanjut, hasil penelitian ini menegaskan bahwa model AI berbasis Isolation Forest berfungsi optimal sebagai alat bantu pendukung pengambilan keputusan, bukan sebagai pengganti keputusan klinis. Model ini berperan sebagai sistem peringatan dini yang menandai data rekam medis dengan potensi ketidakwajaran agar dapat diverifikasi lebih lanjut oleh tenaga kesehatan. Pendekatan ini selaras dengan rekomendasi Organisasi Kesehatan Dunia (WHO, 2017) dan Carayon et al. (2014) yang menekankan pentingnya sistem pendukung keselamatan pasien berbasis teknologi yang bersifat asistif dan berorientasi pada pengurangan risiko, bukan otomatisasi keputusan klinis sepenuhnya. Dalam hal ini, model anomaly detection membantu mengurangi beban kognitif tenaga kesehatan dengan memfokuskan perhatian pada kasus-kasus berisiko tinggi, sehingga berpotensi menurunkan peluang terjadinya kesalahan klinis akibat kualitas data yang buruk.

Dari perspektif implementasi sistem, penggunaan Python dan pustaka open-source dalam pengembangan model memberikan fleksibilitas tinggi untuk pengembangan lanjutan dan integrasi ke dalam Sistem Informasi Manajemen Rumah Sakit (SIMRS). Pendekatan ini sejalan dengan temuan Sendak et al. (2020) dan Rajpurkar et al. (2022) yang menekankan bahwa keberhasilan penerapan AI di layanan kesehatan sangat bergantung pada kemudahan integrasi, transparansi model, dan interpretabilitas hasil bagi

pengguna klinis. Dengan menempatkan model Isolation Forest sebagai lapisan validasi data dalam SIMRS, sistem ini berpotensi berfungsi sebagai mekanisme pengendalian mutu rekam medis secara berkelanjutan. Oleh karena itu, penelitian ini tidak hanya berkontribusi pada pengembangan metodologi deteksi anomali, tetapi juga memberikan implikasi praktis terhadap penguatan tata kelola data klinis dan peningkatan keselamatan pasien berbasis teknologi AI.

SIMPULAN

Penelitian ini menyimpulkan bahwa penerapan model Artificial Intelligence berbasis *unsupervised anomaly detection* dengan algoritma Isolation Forest efektif dalam mengidentifikasi ketidakwajaran pada data rekam medis pasien tanpa memerlukan pelabelan kesalahan sebelumnya. Model mampu mempelajari pola mayoritas data klinis dan membedakan data normal dari data anomali secara konsisten, sebagaimana ditunjukkan oleh distribusi skor anomali dan visualisasi pemisahan yang jelas antar kelompok data. Temuan utama penelitian ini menegaskan bahwa anomali pada rekam medis lebih sering muncul akibat interaksi multivariat, seperti ketidaksesuaian antara nilai laboratorium dan dosis obat, dibandingkan sekadar nilai ekstrem tunggal. Hal ini mengonfirmasi keunggulan pendekatan AI dibandingkan metode konvensional berbasis ambang batas dalam menangkap kompleksitas dan dinamika data klinis, serta memperkuat potensi model sebagai sistem peringatan dini untuk mendukung pencegahan kesalahan teknis dan klinis dalam pengelolaan rekam medis tanpa menggantikan peran pengambilan keputusan tenaga kesehatan.

Sebagai saran pengembangan lanjutan, penelitian berikutnya perlu melibatkan validasi klinis secara langsung melalui kolaborasi dengan tenaga medis guna menilai relevansi dan implikasi klinis dari setiap data yang terdeteksi sebagai anomali. Integrasi model ke dalam Sistem Informasi Manajemen Rumah Sakit (SIMRS) juga perlu diuji secara operasional untuk mengevaluasi dampaknya terhadap alur kerja, beban kerja tenaga kesehatan, serta potensi terjadinya *alarm fatigue*. Selain itu, pengayaan variabel klinis, peningkatan ukuran dataset, serta eksplorasi metode anomaly detection lain atau pendekatan hibrida dapat dilakukan untuk meningkatkan akurasi dan robustnes model. Dengan dukungan implementasi berbasis Python dan pustaka *open-source*, pengembangan berkelanjutan dan replikasi lintas institusi menjadi sangat memungkinkan, sehingga diharapkan model ini dapat berkontribusi nyata dalam peningkatan kualitas data rekam medis, keselamatan pasien, dan tata kelola layanan kesehatan berbasis data.

DAFTAR PUSTAKA

Bates, D. W., Cullen, D. J., Laird, N., Petersen, L. A., Small, S. D., Servi, D., … Leape, L. L. (1999). Computerized physician order entry and medication error prevention. *Journal of the American Medical Informatics Association, 6*(4), 313–321.

Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA, 319*(13), 1317–1318.

Carayon, P., Schoofs Hundt, A., Karsh, B. T., Gurses, A. P., Alvarado, C. J., Smith, M., & Flatley Brennan, P. (2014). Systems engineering approach to patient safety. *Applied Ergonomics, 45*(1), 14–25.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), 1–58.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Machine learning in cardiovascular medicine. *Journal of the American College of Cardiology, 69*(21), 2657–2664.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In *Proceedings of the IEEE International Conference on Data Mining* (pp. 413–422).

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics, 19*(6), 1236–1246.

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). Artificial intelligence in clinical decision support systems. *Nature Medicine, 28*, 31–38.

Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., … Müller, K. R. (2021). Unsupervised anomaly detection: A survey. *ACM Computing Surveys, 54*(2), 1–36.

Sendak, M. P., D'Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., … Balu, S. (2020). A path for translation of machine learning products into healthcare delivery. *NPJ Digital Medicine, 3*, 1–6.

Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. *IEEE Journal of Biomedical and Health Informatics, 22*(5), 1589–1604.

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*, 44–56.

World Health Organization. (2017). *Patient safety: Making health care safer*. Geneva: World Health Organization.