



Perbandingan Metode IQR vs Isolation Forest dalam Deteksi Outlier pada Data Klaim Asuransi di PT Askrindo Cabang Semarang

Jovita Wayan Ramadhani^{1*}, Susanto²

¹Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang

Jl. Soekarno Hatta, Semarang, telp: 024-6702757, e-mail: jovitawayanramadhani@gmail.com

²Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang

Jl. Soekarno Hatta, Semarang, telp: 024-6702757, e-mail: susanto@usm.ac.id

ARTICLE INFO

History of the article :

Received 10 Januari 2026

Received in revised form 25 Januari 2026

Accepted 27 Januari 2026

Available online 28 Januari 2026

Keywords:

Deteksi Anomali; IQR; Isolation Forest; Klaim Asuransi; Outlier

* Correspondence:

Telepon:

+62 81228817976

E-mail:

jovitawayanramadhani@gmail.com

ABSTRACT

Penelitian ini membandingkan metode Interquartile Range (IQR) dan Isolation Forest dalam mendeteksi outlier pada data klaim asuransi di PT Askrindo Cabang Semarang. Data yang dianalisis mencakup nilai klaim dan lama penyelesaiannya. Hasil analisis univariat dengan pendekatan IQR menunjukkan adanya 9 klaim yang menyimpang dari pola umum, yang terdiri atas 5 penyimpangan pada nilai klaim, 3 pada durasi penyelesaian, serta 1 klaim yang menyimpang pada kedua aspek tersebut. Sementara itu, Isolation Forest dengan pendekatan multivariat dan tingkat contamination sebesar 0,05 mendeteksi 6 klaim anomali yang menunjukkan durasi penyelesaian sangat panjang dan kombinasi variabel yang tidak lazim. Hasil penelitian menunjukkan bahwa IQR efektif untuk mengidentifikasi nilai ekstrem secara cepat, sedangkan Isolation Forest lebih mampu menangkap anomali kompleks antarvariabel. Studi ini memberikan perbandingan empiris berbasis data riil industri asuransi yang menegaskan perbedaan karakteristik kedua pendekatan, sehingga memperjelas dasar metodologis dalam pemilihan metode deteksi outlier sesuai tujuan analisis dan karakteristik data klaim.

1. INTRODUCTION

Dalam berbagai bidang, khususnya di sektor keuangan dan asuransi, keberadaan data yang menyimpang dari pola umum atau outlier menjadi perhatian penting karena dapat memengaruhi ketepatan hasil analisis, estimasi risiko, serta kualitas pengambilan keputusan[1][2]. Dalam praktik industri asuransi, outlier biasanya muncul dalam bentuk klaim dengan nilai yang sangat besar, waktu penyelesaian yang menyimpang dari pola normal, atau pola transaksi yang tidak biasa. Situasi tersebut dapat menjadi petunjuk adanya gangguan pada sistem maupun aktivitas yang

berpotensi mencurigakan [3]. Oleh karena itu, penggunaan metode deteksi outlier yang andal menjadi sangat penting untuk mendukung proses audit, evaluasi kinerja, serta upaya pengendalian dan mitigasi risiko operasional[4].

Dalam konteks industri penjaminan kredit, PT Asuransi Kredit Indonesia (Askrindo) Cabang Semarang menangani volume klaim dalam jumlah besar dengan karakteristik data yang cukup beragam dan kompleks. Situasi ini membuat kemungkinan munculnya klaim yang menyimpang dari pola umum, baik dari sisi nilai klaim maupun lama penyelesaian, menjadi semakin besar. Ketidaksesuaian tersebut dapat memengaruhi efisiensi pelayanan, khususnya dalam pemenuhan Service Level Agreement (SLA), serta berdampak pada ketepatan pencatatan dan pengendalian risiko perusahaan.[5].

Berbagai pendekatan telah dikembangkan untuk mendeteksi outlier. Salah satu metode statistik yang cukup sering digunakan adalah Interquartile Range (IQR), yaitu teknik non-parametrik yang sederhana, tidak bergantung pada asumsi distribusi data, dan mampu mengidentifikasi nilai ekstrem secara cepat [6][7]. Namun, karena umumnya diterapkan secara univariat, metode ini memiliki keterbatasan dalam mengenali pola anomali yang melibatkan hubungan antar variabel. Sebagai alternatif, pendekatan berbasis machine learning seperti Isolation Forest dirancang untuk menangani data multivariat dan mendeteksi anomali berdasarkan pola kombinasi fitur tanpa memerlukan label data[3]. Sejumlah penelitian menunjukkan bahwa Isolation Forest memiliki kinerja yang baik dalam berbagai bidang, termasuk sektor keuangan dan keamanan jaringan. Meskipun demikian, metode ini tetap memiliki keterbatasan sehingga penerapannya perlu disesuaikan dengan konteks permasalahan yang dihadapi.

Sejumlah penelitian sebelumnya telah membahas metode deteksi outlier dalam pengolahan data berskala besar dan kompleks. Dalam kajian metode statistik, Dash et al.[8] menunjukkan bahwa penggabungan metode Interquartile Range (IQR) dengan proses validasi musiman mampu meningkatkan kemampuan deteksi outlier sekaligus memperbaiki akurasi prediksi, terutama pada dataset dengan jumlah data yang terbatas. Seiring perkembangan penelitian, metode IQR kemudian banyak dikombinasikan dengan pendekatan machine learning untuk memperkuat kinerja deteksi. Velmurugan et al. [6] serta Sari dan Suharjo[9], misalnya, mengintegrasikan IQR dengan algoritma seperti Local Outlier Factor, Nearest Neighbor, dan teknik reinforcement learning guna meningkatkan performa deteksi pada data asuransi berskala besar yang memiliki keterbatasan label[10]. Sementara itu, Ok et al. [11] menekankan pentingnya penggunaan metrik evaluasi yang lebih tahan terhadap pengaruh outlier, seperti AUCPR dan ROC, dalam menilai kinerja sistem deteksi. Di sisi lain, pendekatan machine learning seperti Isolation Forest (IF) banyak digunakan karena dinilai efisien dalam mendeteksi anomali multivariat tanpa bergantung pada asumsi distribusi data. Metode ini bekerja dengan membangun sejumlah pohon keputusan secara acak dan mengamati seberapa cepat suatu data dapat diisolasi. Semakin cepat sebuah data terpisah dari kumpulan lainnya, semakin besar kemungkinan data tersebut merupakan anomali[12]. Beberapa penelitian menunjukkan potensi besar dari pendekatan ini. Sri Lakshmi et al. [13], misalnya, menemukan bahwa IF memiliki kinerja yang menjanjikan dalam lingkungan Software-Defined Networking, meskipun masih diperlukan penyesuaian untuk meningkatkan metrik evaluasi seperti precision dan recall. Selanjutnya, Shao et al. [14] mengembangkan Cluster-based Improved Isolation Forest (CIIF) dengan mengombinasikan IF dan k-means clustering, yang terbukti mampu meningkatkan stabilitas serta akurasi deteksi anomali hingga rata-rata 7% dibandingkan IF standar. Selain itu, Isolation Forest juga telah diterapkan secara luas di berbagai bidangnya, seperti metrologi dan federated learning, terutama dalam konteks yang menuntut efisiensi tinggi dan perlindungan privasi data [7].

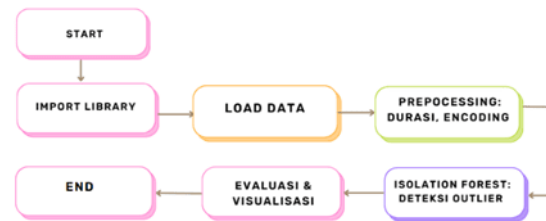
Meskipun kajian mengenai deteksi outlier telah banyak dilakukan, penelitian yang secara langsung membandingkan pendekatan statistik dan machine learning pada data klaim asuransi di

tingkat operasional cabang, khususnya pada perusahaan penjaminan seperti PT Askrindo, masih relatif terbatas. Dengan mempertimbangkan celah penelitian tersebut, studi ini bertujuan membandingkan metode Interquartile Range (IQR) dan Isolation Forest dalam mendeteksi outlier pada data klaim asuransi di PT Askrindo Cabang Semarang, serta menganalisis perbedaan karakteristik hasil deteksi yang dihasilkan oleh kedua pendekatan.

2. RESEARCH METHODS



Gambar 1. Tahapan Penelitian IQR



Gambar 2. Tahapan Penelitian Isolation Forest

a) Identifikasi Masalah

Deteksi outlier pada data klaim asuransi berperan penting dalam mengidentifikasi klaim yang berada di luar pola umum karena dapat menunjukkan adanya ketidakefisienan operasional maupun potensi risiko. Meski demikian, kajian yang membandingkan efektivitas metode statistik seperti Interquartile Range (IQR) dan pendekatan machine learning seperti Isolation Forest pada data klaim asuransi masih terbatas. Berdasarkan hal tersebut, penelitian ini dilakukan untuk membandingkan kedua metode guna menilai efektivitas serta karakteristik hasil deteksi outlier yang dihasilkan.

b) Pengumpulan Data

Penelitian ini menggunakan data historis klaim asuransi dari PT Askrindo Cabang Semarang yang mencakup nilai klaim dan durasi penyelesaian klaim. Seluruh data telah dianonimkan untuk menjaga kerahasiaan tanpa mengubah pola dasar yang relevan untuk analisis.

c) Preprocessing

Tahap preprocessing meliputi validasi data, penanganan nilai hilang, transformasi tanggal menjadi durasi klaim, serta pengodean variabel kategorikal. Mengingat ukuran dataset yang relatif terbatas, pemeriksaan data dilakukan secara menyeluruh untuk memastikan seluruh entri lengkap dan bebas dari nilai kosong sebelum tahap analisis. Untuk penerapan Isolation Forest, variabel kategorikal dikonversi ke bentuk numerik menggunakan teknik label encoding agar dapat diproses oleh model.

d) Metode Interquartile Range (IQR)

Metode Interquartile Range (IQR) digunakan untuk mendeteksi outlier secara univariat dengan memanfaatkan distribusi kuartil data. Nilai IQR dihitung sebagai selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1), yang selanjutnya digunakan untuk menetapkan batas bawah dan batas atas dalam penentuan outlier. Observasi yang berada di luar rentang tersebut dikategorikan sebagai outlier. Pada penelitian ini, metode IQR digunakan pada variabel nilai

klaim dan durasi penyelesaian untuk mendeteksi nilai ekstrem secara cepat dan mudah dipahami.

Jangkauan Interquartile Range (IQR):

$$IQR = Q_3 - Q_1 \quad (1)$$

Persamaan (1) menghitung IQR sebagai selisih kuartil ketiga (Q_3) dan pertama (Q_1), untuk mengukur sebar data dan mendeteksi outlier.

Batas Bawah (Lower Bound):

$$Upper\ Bound = Q_1 - 1.5 \times IQR \quad (2)$$

Persamaan(2) menentukan batas bawah yang dianggap sebagai outlier.

Batas Atas (Upper Bound):

$$Upper\ Bound = Q_3 + 1.5 \times IQR \quad (3)$$

Persamaan (3) menentukan batas atas yang dianggap sebagai outlier atas.

e) Metode Isolation Forest

Isolation Forest merupakan algoritma unsupervised learning yang mendeteksi anomali berdasarkan seberapa mudah suatu observasi dapat dipisahkan dalam kumpulan pohon keputusan yang dibentuk secara acak. Observasi yang dapat diisolasi lebih cepat memiliki panjang jalur isolasi yang lebih pendek dan cenderung dikategorikan sebagai outlier. Dalam penelitian ini, Isolation Forest digunakan untuk mengidentifikasi klaim dengan pola yang tidak lazim berdasarkan kombinasi nilai klaim dan durasi penyelesaian. Sebelum proses pemodelan, data distandardisasi menggunakan StandardScaler untuk memastikan setiap variabel berkontribusi secara seimbang. Skor anomali kemudian dihitung berdasarkan rata-rata panjang jalur isolasi yang telah dinormalisasi, di mana nilai skor yang mendekati 1 menunjukkan tingkat keanomalien yang lebih tinggi. Parameter contamination ditetapkan sebesar 0,05, yang mengindikasikan bahwa sekitar 5% data diasumsikan sebagai outlier. Nilai ini dipilih sebagai pendekatan yang moderat dan umum digunakan dalam analisis unsupervised, khususnya ketika proporsi anomali tidak diketahui secara pasti, sehingga hasil deteksi tidak bersifat terlalu agresif maupun terlalu konservatif.

Standardisasi Data (StandardScaler):

$$z_i = \frac{x_i - \mu}{\sigma} \quad (4)$$

Persamaan (4) menstandarisasi data menjadi rata-rata nol dan deviasi standar satu. Standardisasi pada nilai klaim dan durasi resolusi memastikan kedua variabel berkontribusi seimbang dalam Isolation Forest, mencegah bias terhadap variabel berskala besar.

Panjang Jalur Isolasi (Path Length):

$$h(x) = \text{jumlah pemisah yang diperlukan untuk mengisolasi } x \quad (5)$$

Persamaan (5) Panjang jalur $h(x)$ menunjukkan seberapa cepat sebuah observasi terisolasi dalam isolation tree. Data anomali terisolasi lebih cepat sehingga $h(x)$ lebih pendek dibanding data normal sehingga panjang jalur ini digunakan untuk menilai keanomalien klaim.

Rata-rata Panjang Jalur pada Forest:

$$E(h(x)) = \frac{1}{t} \sum_{i=1}^t h_1(x) \quad (6)$$

Persamaan (6) menghitung rata-rata panjang jalur isolasi suatu observasi di seluruh pohon Isolation Forest. Semakin kecil nilai rata-rata, semakin tinggi kemungkinan klaim tersebut merupakan anomali.

Normalisasi Panjang Jalur:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \quad (7)$$

Dengan:

$$H(n) = \ln(n) + \gamma$$

Persamaan (7) fungsi $c(n)$ menormalisasi panjang jalur isolasi sesuai ukuran sampel, merepresentasikan ekspektasi panjang pada binary search tree tak seimbang. Normalisasi ini membuat skor anomali bisa dibandingkan secara adil meski jumlah data berbeda.

Skor Anomali Isolation Forest:

$$s(x, n) = 2 \frac{E(h(x))}{c(n)} \quad (8)$$

Persamaan (8) menghasilkan skor anomali antara 0–1. Semakin mendekati 1, semakin tinggi kemungkinan klaim menjadi outlier, artinya nilai klaim atau durasi penyelesaiannya menyimpang dari mayoritas.

Ambang Batas Outlier:

$$Outlier = \begin{cases} 1, & \text{jika } s(x) \geq \tau \\ 0, & \text{jika } s(x) < \tau \end{cases} \quad (9)$$

Persamaan (9) Ambang batas τ ditentukan dari parameter contamination, yaitu proporsi data yang diasumsikan outlier. Dengan contamination 5%, model mengklasifikasikan 5% data dengan skor anomali tertinggi sebagai klaim anomali, sebagai estimasi awal pada analisis unsupervised.

f) Pengaturan Eksperimen dan Evaluasi

Model Isolation Forest diaplikasikan menggunakan dua variabel utama, yakni nilai klaim dan durasi penyelesaian klaim. Sementara itu, metode IQR digunakan secara univariat pada masing-masing variabel. Hasil deteksi dari kedua pendekatan selanjutnya dibandingkan dengan melihat jumlah outlier yang teridentifikasi, karakteristik klaim yang terdeteksi, serta perbedaan pola anomali yang muncul. Analisis visualisasi digunakan sebagai pendukung untuk memperjelas sebaran data dan membandingkan kinerja pendekatan univariat dan multivariat dalam mendeteksi outlier.

3. RESULTS

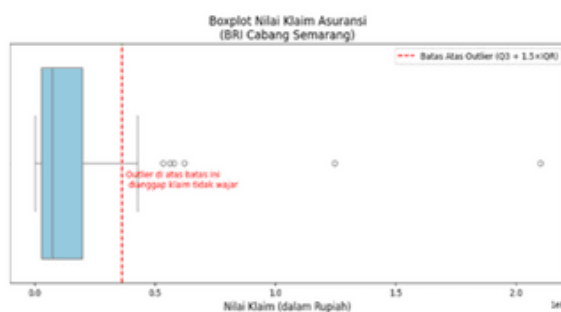
Metode Interquartile Range (IQR) digunakan untuk mendeteksi outlier pada variabel nilai klaim dan durasi penyelesaian klaim (dalam hari) dengan menetapkan batas bawah dan batas atas

berdasarkan selisih kuartil pertama (Q1) dan kuartil ketiga (Q3). Berdasarkan Tabel 1, ditemukan 9 klaim yang tergolong sebagai outlier. Dari jumlah tersebut, 5 klaim memiliki nilai yang tergolong ekstrem, 3 klaim menunjukkan durasi penyelesaian yang tidak lazim, dan 1 klaim menyimpang pada kedua variabel. Outlier nilai klaim ditunjukkan oleh nominal yang jauh lebih tinggi dibandingkan mayoritas data, seperti pada Unit xxxx dengan nilai klaim Rp62.251.248,5 dan durasi 157 hari, sedangkan outlier durasi dicirikan oleh waktu penyelesaian yang sangat panjang, salah satunya mencapai 593 hari pada Unit xxxx. Selain itu, satu klaim pada Unit xxxx dengan nilai Rp33.305.539,4 dan durasi 80 hari teridentifikasi sebagai outlier ganda karena kedua variabel berada pada tingkat ekstrem. Hasil ini menunjukkan bahwa metode IQR efektif dalam mendeteksi penyimpangan data secara univariat.

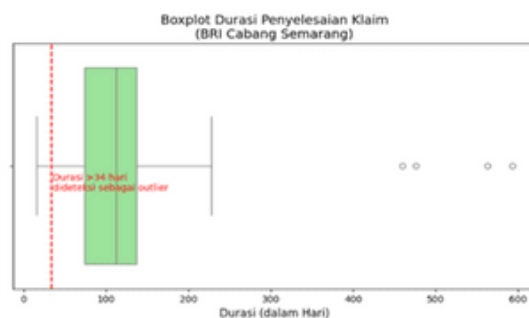
Tabel 1.
Ringkasan Outlier Metode IQR

No.	Cabang/Area	Nilai Klaim	Durasi Hari	Kategori Outlier
1.	Unit xxxx	Rp33.305.539,4	80	Nilai Klaim
2.	Unit xxxx	Rp13.015.807,0	134	Nilai Klaim
3.	Unit xxxx	Rp1.645.242,9	132	Nilai Klaim
4.	Unit xxxx	Rp21.000.000,0	72	Nilai Klaim
5.	Unit xxxx	Rp62.251.248,5	157	Nilai Klaim
6.	Unit xxxx	Rp173.688,9	30	Durasi Hari
7.	Unit xxxx	Rp1.487.734,5	60	Durasi Hari
8.	Unit xxxx	Rp6.179.506,2	593	Durasi Hari
9.	Unit xxxx	Rp33.305.539,4	80	Durasi Hari

Secara umum, temuan ini menunjukkan bahwa klaim dengan nilai yang lebih besar cenderung membutuhkan waktu penyelesaian yang lebih lama. Hal ini mengindikasikan adanya keterkaitan antara besarnya nilai klaim dan tingkat kompleksitas proses penanganan klaim.



Gambar 3. Boxplot Nilai Klaim



Gambar 4. Boxplot Durasi Penyelesaian Klaim

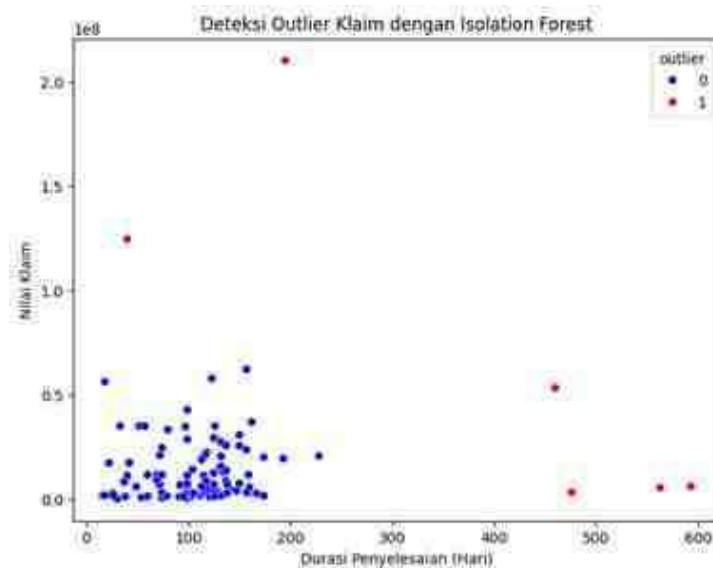
Pada gambar 3 menunjukkan boxplot nilai klaim asuransi di Cabang Semarang. Garis merah putus-putus adalah batas atas outlier menurut metode IQR ($Q3 + 1.5 \times IQR$). Titik-titik di sebelah kanan garis tersebut mewakili klaim klaim yang dianggap outlier karena nilainya jauh lebih tinggi dari data mayoritas. Sedangkan pada Gambar 4 menampilkan boxplot durasi penyelesaian klaim. Garis merah menunjukkan batas atas outlier yaitu 34 hari. Titik-titik sebelah kanan garis tersebut adalah klaim yang memakan waktu lebih dari 34 hari dan dianggap outlier.

Selain metode IQR, Isolation Forest (IF) digunakan sebagai pendekatan pembandingan untuk mengidentifikasi klaim yang menyimpang. Metode ini mengenali anomali melalui proses isolasi observasi dalam struktur pohon acak tanpa memerlukan asumsi distribusi data. Dalam penerapannya, model IF dilatih menggunakan dua variabel utama, yakni nilai klaim dan durasi penyelesaian klaim, setelah data terlebih dahulu diproses dan distandardisasi. Parameter contamination ditetapkan sebesar 0,05, yang mengindikasikan bahwa sekitar 5% data diasumsikan sebagai anomali. Nilai ini dipilih sebagai pendekatan yang moderat dan lazim digunakan dalam analisis unsupervised ketika proporsi outlier tidak diketahui secara pasti, sehingga hasil deteksi tetap representatif tanpa bersifat terlalu agresif maupun terlalu konservatif.

Tabel 2.
Ringkasan Outlier Metode Isolation Forest

Nilai Klaim	Durasi Hari	Alasan Anomali
Rp210.000.000	195	Nilai klaim sangat ekstrem
Rp124.483.401	40	Klaim besar, waktu relatif singkat
Rp53.319.553	460	Klaim besar dan durasi sangat lama
Rp6.179.506	593	Durasi ekstrem
Rp5.554.024	563	Durasi ekstrem
Rp3.301.207	476	Durasi ekstrem

Pada tabel 2. Berdasarkan metode Isolation Forest, beberapa klaim diidentifikasi sebagai outlier. Hasilnya menunjukkan bahwa klaim-klaim ini memiliki nilai klaim dan /atau durasi penyelesaian yang berbeda secara signifikan dari sebagian besar data lainnya. Beberapa klaim ditandai sebagai outlier karena memiliki nilai klaim yang sangat tinggi, sementara yang lain memiliki durasi penyelesaian yang jauh lebih lama dari biasanya. Hal ini menunjukkan bahwa deteksi outlier dipengaruhi tidak hanya oleh besarnya nilai klaim tetapi juga oleh lamanya proses penyelesaian klaim. Oleh karena itu, tabel outlier menyoroti klaim – klaim yang perlu diselidiki lebih lanjut karena memiliki pola yang tidak biasa dibandingkan dengan klaim - klaim normal.



Gambar 5. Visualisasi Outlier dengan Isolation Forest (Pairplot)

Pada Gambar 5 menampilkan hasil visualisasi deteksi outlier menggunakan metode Isolation Forest dengan mempertimbangkan nilai klaim dan durasi penyelesaian secara bersamaan. Klaim yang berada dalam kondisi normal cenderung terkonsentrasi pada rentang nilai dan durasi yang relatif moderat, sedangkan klaim yang teridentifikasi sebagai outlier tampak terpisah dari kelompok utama. Pemisahan tersebut umumnya dipengaruhi oleh nilai klaim yang sangat tinggi, durasi penyelesaian yang jauh lebih lama, atau kombinasi dari kedua karakteristik tersebut. Pola ini mengindikasikan bahwa Isolation Forest mampu mengenali dan mengisolasi klaim dengan karakteristik ekstrem secara multivariat, sehingga mendukung hasil deteksi outlier yang dihasilkan oleh model.

4. DISCUSSION

Hasil analisis menunjukkan bahwa pola deteksi outlier yang dihasilkan oleh metode Interquartile Range (IQR) dan Isolation Forest tidak sepenuhnya sama pada data klaim asuransi PT Askrindo Cabang Semarang. Metode IQR yang diterapkan secara univariat mampu mengidentifikasi nilai ekstrem pada nilai klaim dan durasi penyelesaian secara terpisah. Hal ini sejalan dengan temuan penelitian sebelumnya yang menempatkan IQR sebagai metode yang sederhana dan efektif untuk deteksi awal. Namun, karena bersifat univariat, IQR memiliki keterbatasan dalam menangkap anomali yang melibatkan keterkaitan antarvariabel. Berbeda dari IQR, Isolation Forest melakukan deteksi outlier dengan melihat nilai klaim dan durasi penyelesaian secara bersamaan. Pendekatan machine learning ini lebih efektif dalam mengenali pola anomali multivariat dan interaksi antarvariabel. Klaim yang teridentifikasi sebagai outlier umumnya ditandai oleh kombinasi nilai klaim yang tinggi dan durasi penyelesaian yang panjang, yang juga terlihat jelas pada hasil visualisasi. Temuan ini konsisten dengan penelitian sebelumnya yang menyebutkan bahwa Isolation Forest unggul dalam mendeteksi anomali kompleks tanpa bergantung pada asumsi distribusi data.

Dalam praktiknya, IQR lebih cocok digunakan sebagai alat awal untuk menandai klaim dengan nilai atau durasi yang tidak lazim. Untuk analisis yang lebih mendalam, Isolation Forest lebih sesuai karena mampu mengidentifikasi pola klaim yang menyimpang secara lebih kompleks.

Dengan memadukan kedua metode tersebut, pengawasan terhadap proses klaim dapat dilakukan secara lebih menyeluruh dan mendukung pengambilan keputusan yang lebih tepat.

5. CONCLUSIONS AND RECOMMENDATIONS

Hasil analisis menunjukkan bahwa IQR dan Isolation Forest sama-sama dapat mengenali outlier, namun melalui pendekatan yang tidak sama. IQR lebih efektif untuk menandai nilai ekstrem secara cepat, sementara Isolation Forest lebih sensitif dalam mengenali anomali yang muncul dari kombinasi nilai klaim dan durasi penyelesaian. Oleh karena itu, kedua metode ini lebih optimal digunakan secara berurutan, dengan IQR sebagai tahap awal dan Isolation Forest untuk analisis lanjutan. Kontribusi utama penelitian ini adalah penyajian perbandingan empiris berbasis data riil industri asuransi yang memberikan dasar metodologis yang jelas dalam pemilihan metode deteksi outlier sesuai tujuan analisis.

6. REFERENCES

- [1] D. L. Whaley, "The Interquartile Range: Theory and Estimation," *Electron. Theses Diss.*, p. 54, 2005, [Online]. Available: <http://dc.etsu.edu/etd%0Ahttp://dc.etsu.edu/etd>
- [2] D. Herurkar, S. Palacio, A. Anwar, J. Hees, and A. Dengel, "Fin-Fed-OD: Federated Outlier Detection on Financial Tabular Data," 2024, [Online]. Available: <http://arxiv.org/abs/2404.14933>
- [3] S. Buschjäger, P. J. Honysz, and K. Morik, "Randomized outlier detection with trees," *Int. J. Data Sci. Anal.*, vol. 13, no. 2, pp. 91–104, 2022, doi: 10.1007/s41060-020-00238-w.
- [4] A. Capozzi *et al.*, "FlowSeries: Anomaly Detection in Financial Transaction Flows," *Stud. Comput. Intell.*, vol. 1189 SCI, pp. 29–40, 2025, doi: 10.1007/978-3-031-82435-7_3.
- [5] P. K. Dongre, V. Patel, U. Bhoi, and N. N. Maltare, "An outlier detection framework for Air Quality Index prediction using linear and ensemble models," *Decis. Anal. J.*, vol. 14, no. July 2024, p. 100546, 2025, doi: 10.1016/j.dajour.2025.100546.
- [6] Dr. Velmurugan. K, Mr. K. Pazhanivel, Divyasree. R, Gowtham. E, and Guruharan. S, "Data Driven Analysis of Insurance Claims Using Machine Learning Algorithm," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 269–273, 2023, doi: 10.48175/ijarsct-9689.
- [7] I. Forest and I. Forest, "Isolation Forest算法实现详解（点击阅读原文有惊喜哦），" vol. 9818, no. 1, pp. 1–15, 2021.
- [8] R. Items, W. Rose, W. Rose, T. If, and W. Rose, "Encyclopedia of business analytics and optimization," *Choice Rev. Online*, vol. 52, no. 02, pp. 52-0606-52–0606, 2014, doi: 10.5860/choice.52-0606.
- [9] P. O. C. Sari and S. Suharjito, "Outlier Detection in Inpatient Claims Using DBSCAN and K-Means," *J. Tek. Inform.*, vol. 15, no. 1, pp. 1–10, 2022, doi: 10.15408/jti.v15i1.25682.
- [10] B. Avanzi, M. Lavender, G. Taylor, and B. Wong, "Detection and treatment of outliers for

- multivariate robust loss reserving,” *Ann. Actuar. Sci.*, vol. 18, no. 1, pp. 102–125, 2024, doi: 10.1017/S1748499523000155.
- [11] M. Ok, S. Klüttermann, and E. Müller, “Exploring the Impact of Outlier Variability on Anomaly Detection Evaluation Metrics,” 2024, [Online]. Available: <http://arxiv.org/abs/2409.15986>
- [12] R. Selina, Y. Yanti, and A. Suryono, “Analisis Yuridis Tanggung Jawab Perusahaan Asuransi Terhadap Asuransi Jiwa Kredit Rahma Selina Yustika Yanti Arief Suryono K / Pdt / 2019). Penolakan pembayaran klaim meninggal dunia dilakukan oleh pihak PT . Asuransi Jiwa Inhealth Indonesia kepada pemeg,” no. 1, 2024.
- [13] M. Sri Lakshmi, G. Rajavikram, V. Dattatreya, B. Swarna Jyothi, S. Patil, and M. Bhavsingh, “Evaluating the Isolation Forest Method for Anomaly Detection in Software-Defined Networking Security,” *J. Electr. Syst.*, vol. 19, no. 4, pp. 279–297, 2023, doi: 10.52783/jes.639.
- [14] C. Shao, X. Du, J. Yu, and J. Chen, “Cluster-Based Improved Isolation Forest,” *Entropy*, vol. 24, no. 5, 2022, doi: 10.3390/e24050611.