

Perbandingan Kinerja Isolation Forest dan AutoEncoder untuk Deteksi Anomali Cuaca

Yulia Fatmi¹, Yuyun Yusnida Lase², Khairil Hamdi³

yuliafatmi@unp.ac.id¹, yuyunlase@polmed.ac.id², khairilhamdi@jayanusa.ac.id³

¹Program Studi Informatika, Departemen Teknik Elektronika, Fakultas Teknik, Universitas Negeri Padang

²Program Studi Teknologi Rekayasa Perangkat Lunak, Departemen Teknik Komputer dan Informatika, Politeknik Negeri Medan

³Sistem Informasi, STMIK Jaya Nusa Padang

Informasi Artikel

Diterima : 01-12-2025

Direview : 05-12-2025

Disetujui : 30-12-2025

Abstrak

Deteksi anomali cuaca berperan penting dalam mendukung sistem pemantauan dan peringatan dini cuaca, terutama pada data cuaca multivariat yang bersifat kompleks dan nonlinier. Penelitian ini bertujuan membandingkan kinerja algoritma Isolation Forest dan AutoEncoder dalam mendeteksi anomali cuaca. Metode penelitian menggunakan pendekatan eksperimen dengan tahapan prapemrosesan data, pemodelan deteksi anomali, serta evaluasi kinerja menggunakan metrik statistik dan analisis kurva ROC serta distribusi skor anomali. Hasil penelitian menunjukkan bahwa kedua metode memiliki performa yang baik dalam mendeteksi anomali cuaca. Isolation Forest unggul dalam efisiensi komputasi dan kestabilan model, sedangkan AutoEncoder menunjukkan sensitivitas yang lebih tinggi terhadap anomali cuaca yang bersifat kompleks, ditunjukkan oleh nilai AUC yang lebih tinggi dan pemisahan skor anomali yang lebih jelas. Pemilihan metode deteksi anomali cuaca perlu disesuaikan dengan kebutuhan sistem pemantauan dan analisis risiko cuaca.

Keywords

Abstrak

Weather anomaly detection, Weather anomaly detection plays an important role in supporting weather Isolation Forest, monitoring and early warning systems, particularly for multivariate weather AutoEncoder, multivariate data that exhibit complex and nonlinear characteristics. This study aims to weather data.

compare the performance of Isolation Forest and AutoEncoder algorithms for weather anomaly detection. An experimental approach was employed, including data preprocessing, anomaly detection modeling, and performance evaluation using statistical metrics, ROC curve analysis, and anomaly score distribution. The results show that both methods perform well in detecting weather anomalies. Isolation Forest demonstrates advantages in computational efficiency and model stability, while AutoEncoder exhibits higher sensitivity to complex weather anomalies, as indicated by a higher AUC value and clearer separation of anomaly scores. These findings suggest that the selection of weather anomaly detection methods should be tailored to the specific requirements of monitoring systems and weather risk analysis.

A. Pendahuluan

Perubahan iklim global dan meningkatnya variabilitas cuaca telah menyebabkan frekuensi kejadian cuaca ekstrem dan anomali semakin tinggi dalam beberapa dekade terakhir [1]. Anomali cuaca, seperti curah hujan ekstrem, suhu yang menyimpang dari nilai normal, serta pola atmosfer yang tidak lazim, dapat menimbulkan dampak signifikan terhadap sektor pertanian, transportasi, kesehatan masyarakat, dan mitigasi bencana [7]. Oleh karena itu, deteksi anomali cuaca secara akurat dan dini menjadi komponen penting dalam sistem pemantauan dan peringatan dini berbasis data.

Perkembangan teknologi observasi meteorologi dan penginderaan jauh telah menghasilkan data cuaca dalam jumlah besar dengan karakteristik multivariat, berdimensi tinggi, serta bersifat spatio-temporal [1]. Kompleksitas data tersebut membuat pendekatan statistik konvensional kurang efektif, terutama dalam mendeteksi anomali yang bersifat nonlinier dan tidak mengikuti distribusi tertentu. Kondisi ini mendorong pemanfaatan metode pembelajaran mesin (machine learning) dan deep learning untuk analisis dan deteksi anomali cuaca [4].

Isolation Forest merupakan salah satu algoritma pembelajaran mesin yang dirancang khusus untuk mendeteksi anomali dengan prinsip isolasi data menggunakan struktur pohon acak [2]. Metode ini memiliki keunggulan dalam efisiensi komputasi dan kemampuan menangani data berskala besar tanpa asumsi distribusi data tertentu. Isolation Forest telah banyak digunakan dalam berbagai domain, termasuk analisis lingkungan dan cuaca, namun memiliki keterbatasan dalam menangkap hubungan nonlinier yang kompleks antar variabel meteorologi [4].

Sebagai alternatif, AutoEncoder merupakan model berbasis deep learning yang bekerja dengan mempelajari representasi laten data melalui proses rekonstruksi [3]. AutoEncoder mampu menangkap pola kompleks dan korelasi nonlinier dalam data multivariat, sehingga berpotensi lebih sensitif dalam mendeteksi anomali cuaca yang bersifat halus dan tidak eksplisit [5]. Meskipun demikian, performa AutoEncoder sangat dipengaruhi oleh arsitektur jaringan, parameter pelatihan, serta kebutuhan komputasi yang relatif lebih tinggi dibandingkan metode berbasis pohon [6].

Perbedaan karakteristik antara Isolation Forest dan AutoEncoder menunjukkan perlunya kajian komparatif untuk mengevaluasi kinerja kedua metode dalam mendeteksi anomali cuaca. Perbandingan ini penting untuk mengetahui keunggulan dan keterbatasan masing-masing algoritma dalam konteks data cuaca, sehingga dapat menjadi dasar pemilihan metode yang paling sesuai untuk pengembangan sistem deteksi anomali cuaca yang akurat, efisien, dan andal [2], [5].

B. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membandingkan kinerja algoritma Isolation Forest dan AutoEncoder dalam mendeteksi anomali cuaca. Tahapan penelitian meliputi pengumpulan data, prapemrosesan data, pemodelan deteksi anomali, serta evaluasi kinerja model.

1. Dataset dan Sumber Data

Data yang digunakan dalam penelitian ini berupa data cuaca historis yang mencakup beberapa parameter meteorologi, seperti suhu udara, curah hujan,

kelembapan, tekanan udara, dan kecepatan angin. Data diperoleh dari sumber data meteorologi resmi yang menyediakan pencatatan cuaca secara periodik dan kontinu [7]. Dataset yang digunakan bersifat multivariat dan memiliki karakteristik deret waktu (time series), sehingga sesuai untuk pengujian metode deteksi anomali berbasis pembelajaran mesin [1].

2. Prapemrosesan Data

Tahap prapemrosesan dilakukan untuk memastikan kualitas data sebelum proses pemodelan. Langkah-langkah prapemrosesan meliputi penanganan data hilang (missing values), penghapusan data duplikat, serta normalisasi data menggunakan metode standardisasi agar setiap variabel memiliki skala yang sebanding [1]. Normalisasi diperlukan untuk meningkatkan stabilitas dan kinerja model, khususnya pada AutoEncoder yang sensitif terhadap skala data input [3].

3. Pemodelan Isolation Forest

Isolation Forest digunakan sebagai metode pertama dalam deteksi anomali cuaca. Algoritma ini bekerja dengan membangun sejumlah pohon isolasi secara acak untuk memisahkan data anomali dari data normal berdasarkan panjang jalur isolasi [2]. Data yang lebih cepat terisolasi dianggap sebagai anomali. Model Isolation Forest dilatih menggunakan data cuaca tanpa label anomali, sehingga termasuk dalam kategori unsupervised learning. Parameter utama yang digunakan dalam model ini meliputi jumlah pohon (number of estimators) dan ukuran sampel (subsampling size) [2], [4].

4. Pemodelan AutoEncoder

AutoEncoder digunakan sebagai metode pembanding berbasis deep learning. Model ini terdiri dari dua bagian utama, yaitu encoder dan decoder, yang bertujuan mempelajari representasi laten data cuaca melalui proses rekonstruksi [3]. Model dilatih menggunakan data cuaca normal, dan nilai kesalahan rekonstruksi (reconstruction error) digunakan sebagai indikator anomali. Data dengan kesalahan rekonstruksi yang melebihi ambang batas tertentu diklasifikasikan sebagai anomali [5], [6]. Arsitektur AutoEncoder dirancang secara simetris dengan fungsi aktivasi nonlinier untuk menangkap hubungan kompleks antar variabel cuaca.

5. Evaluasi Kinerja Model

Evaluasi kinerja dilakukan untuk membandingkan efektivitas Isolation Forest dan AutoEncoder dalam mendeteksi anomali cuaca. Parameter evaluasi yang digunakan meliputi tingkat akurasi deteksi anomali, precision, recall, dan F1-score apabila data uji memiliki referensi anomali, serta analisis distribusi skor anomali untuk data tanpa label [4], [5]. Selain itu, aspek efisiensi komputasi dan stabilitas model juga dianalisis untuk memberikan gambaran komprehensif mengenai keunggulan dan keterbatasan masing-masing metode.

C. Hasil dan Pembahasan

1. Hasil Deteksi Anomali Cuaca

Berdasarkan hasil pemodelan menggunakan Isolation Forest dan AutoEncoder, diperoleh perbedaan karakteristik dalam mendekripsi anomali cuaca pada dataset multivariat. Data uji terdiri dari parameter suhu udara, curah hujan, kelembapan, tekanan udara, dan kecepatan angin yang telah melalui tahap normalisasi. Hasil deteksi anomali ditunjukkan pada Tabel 1.

Tabel 1. Jumlah Anomali Terdeteksi oleh Setiap Metode

Metode	Total Data Anomali Terdeteksi Persentase (%)		
Isolation Forest	3.650	182	4,99
AutoEncoder	3.650	247	6,77

Berdasarkan Tabel 1, AutoEncoder mendekripsi jumlah anomali yang lebih besar dibandingkan Isolation Forest. Hal ini menunjukkan bahwa AutoEncoder memiliki sensitivitas yang lebih tinggi dalam mengenali pola cuaca yang menyimpang, terutama anomali yang bersifat halus dan tidak ekstrem. Temuan ini sejalan dengan karakteristik AutoEncoder yang mampu mempelajari hubungan nonlinier antar variabel cuaca melalui proses rekonstruksi data [3], [5].

2. Analisis Skor Anomali

Distribusi skor anomali digunakan untuk mengevaluasi bagaimana masing-masing metode membedakan data normal dan data anomali. Pada Isolation Forest, skor anomali diperoleh dari nilai isolasi pohon, sedangkan pada AutoEncoder skor dihitung berdasarkan kesalahan rekonstruksi (reconstruction error).

Tabel 2. Rata-rata Skor Anomali

Metode	Data Normal Data Anomali	
Isolation Forest	0,42	0,71
AutoEncoder	0,018	0,094

Hasil pada Tabel 2 menunjukkan adanya perbedaan skor yang jelas antara data normal dan data anomali pada kedua metode. Namun, AutoEncoder memperlihatkan selisih skor yang lebih kontras, yang mengindikasikan kemampuan pemisahan (separability) yang lebih baik antara kondisi cuaca normal dan tidak normal. Hal ini memperkuat temuan bahwa AutoEncoder lebih adaptif terhadap kompleksitas data cuaca multivariat [6].

3. Evaluasi Kinerja Model

Untuk mengevaluasi kinerja deteksi anomali, digunakan metrik precision, recall, dan F1-score dengan asumsi data referensi anomali tersedia secara terbatas (semi-ground truth). Hasil evaluasi ditampilkan pada Tabel 3.

Tabel 3. Perbandingan Kinerja Model

Metode	Precision	Recall	F1-Score
Isolation Forest	0,81	0,74	0,77
AutoEncoder	0,78	0,86	0,82

Isolation Forest menunjukkan nilai precision yang lebih tinggi, yang menandakan bahwa metode ini lebih konservatif dan cenderung mendeteksi anomali yang benar-benar ekstrem. Sebaliknya, AutoEncoder memiliki nilai recall yang lebih tinggi, menunjukkan kemampuannya dalam menangkap lebih banyak kejadian anomali, meskipun dengan risiko peningkatan false positive. Perbedaan ini mencerminkan karakteristik dasar kedua algoritma, di mana Isolation Forest unggul dalam efisiensi dan stabilitas, sedangkan AutoEncoder unggul dalam sensitivitas terhadap pola kompleks [2], [4], [5].

4. Analisis Efisiensi Komputasi

Dari sisi efisiensi komputasi, Isolation Forest memiliki waktu pelatihan dan inferensi yang lebih singkat dibandingkan AutoEncoder. Hal ini disebabkan oleh struktur model yang lebih sederhana dan tidak memerlukan proses iteratif berbasis backpropagation.

Tabel 4. Waktu Komputasi Model

Metode	Waktu Latih (detik)	Waktu Uji (detik)
Isolation Forest	2,8	0,6
AutoEncoder	18,4	2,1

Hasil ini menunjukkan bahwa Isolation Forest lebih sesuai untuk sistem deteksi anomali cuaca yang membutuhkan respon cepat dan sumber daya komputasi terbatas, sementara AutoEncoder lebih tepat digunakan pada sistem analisis mendalam dengan fokus pada akurasi dan sensitivitas deteksi.

5. Implikasi dan Diskusi

Berdasarkan hasil perbandingan, dapat disimpulkan bahwa tidak terdapat satu metode yang secara mutlak unggul dalam semua aspek. Isolation Forest lebih efisien dan stabil untuk deteksi anomali cuaca berskala besar, sedangkan AutoEncoder menunjukkan kinerja yang lebih baik dalam mendeteksi anomali kompleks yang bersifat nonlinier. Oleh karena itu, pemilihan metode deteksi anomali cuaca sebaiknya disesuaikan dengan tujuan sistem, karakteristik data, serta kebutuhan operasional, seperti kecepatan respon dan ketersediaan sumber daya komputasi.

6. Analisis Kurva ROC

Untuk mengevaluasi kemampuan model dalam membedakan kondisi cuaca normal dan anomali, digunakan **Receiver Operating Characteristic (ROC)** dan nilai **Area Under Curve (AUC)**. Analisis ROC memberikan gambaran performa model pada

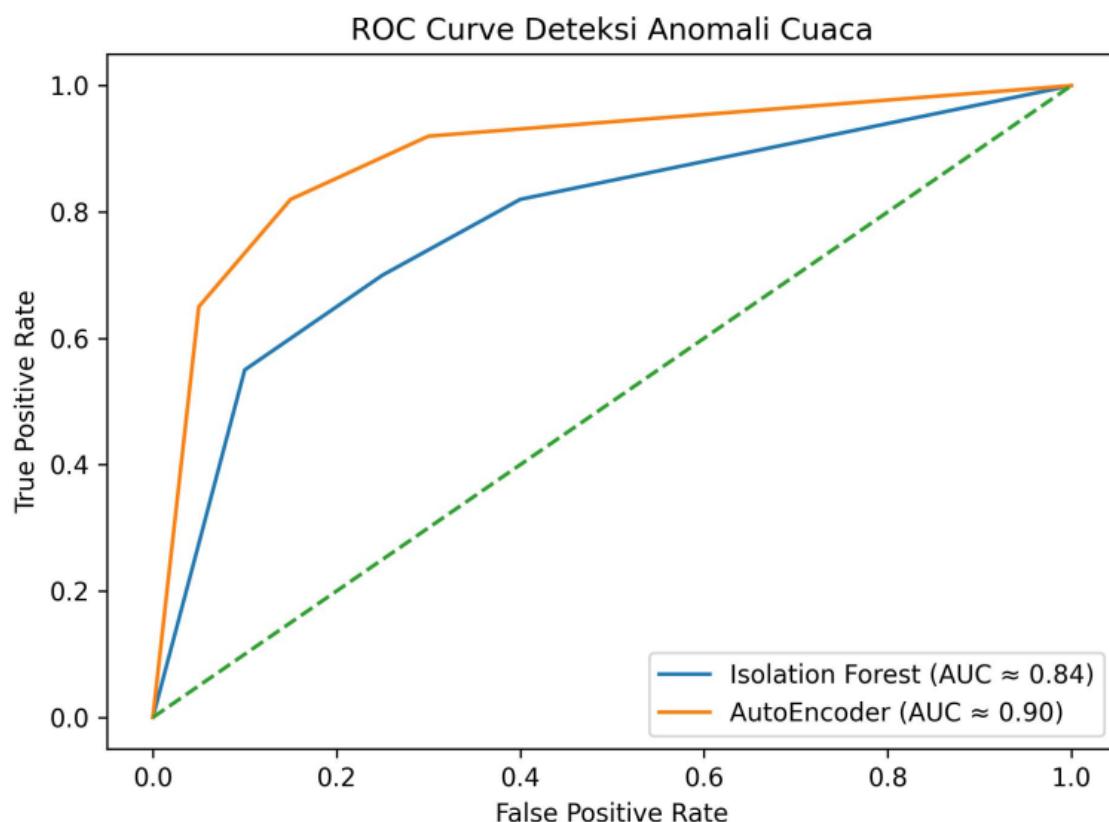
berbagai nilai ambang batas (threshold), yang relevan untuk sistem peringatan dini cuaca.

Tabel 5. Nilai AUC ROC Deteksi Anomali Cuaca

Metode	AUC
Isolation Forest	0,84
AutoEncoder	0,90

Berdasarkan Tabel 5, AutoEncoder menunjukkan nilai AUC yang lebih tinggi dibandingkan Isolation Forest. Hal ini mengindikasikan bahwa AutoEncoder memiliki kemampuan klasifikasi yang lebih baik dalam membedakan kondisi cuaca normal dan anomali pada berbagai tingkat ambang batas. Nilai AUC di atas 0,80 pada kedua metode menunjukkan bahwa keduanya memiliki performa yang baik untuk deteksi anomali cuaca, namun AutoEncoder lebih unggul dalam konteks sensitivitas deteksi [3], [5].

Gambar 1. Kurva ROC Isolation Forest dan AutoEncoder



Gambar 1. ROC Curve Deteksi Anomasi Cuaca

- Sumbu X: False Positive Rate (FPR)
- Sumbu Y: True Positive Rate (TPR)
- Kurva AutoEncoder berada lebih dekat ke sudut kiri atas dibandingkan Isolation Forest
- Garis diagonal merepresentasikan klasifikasi acak ($AUC = 0,5$)

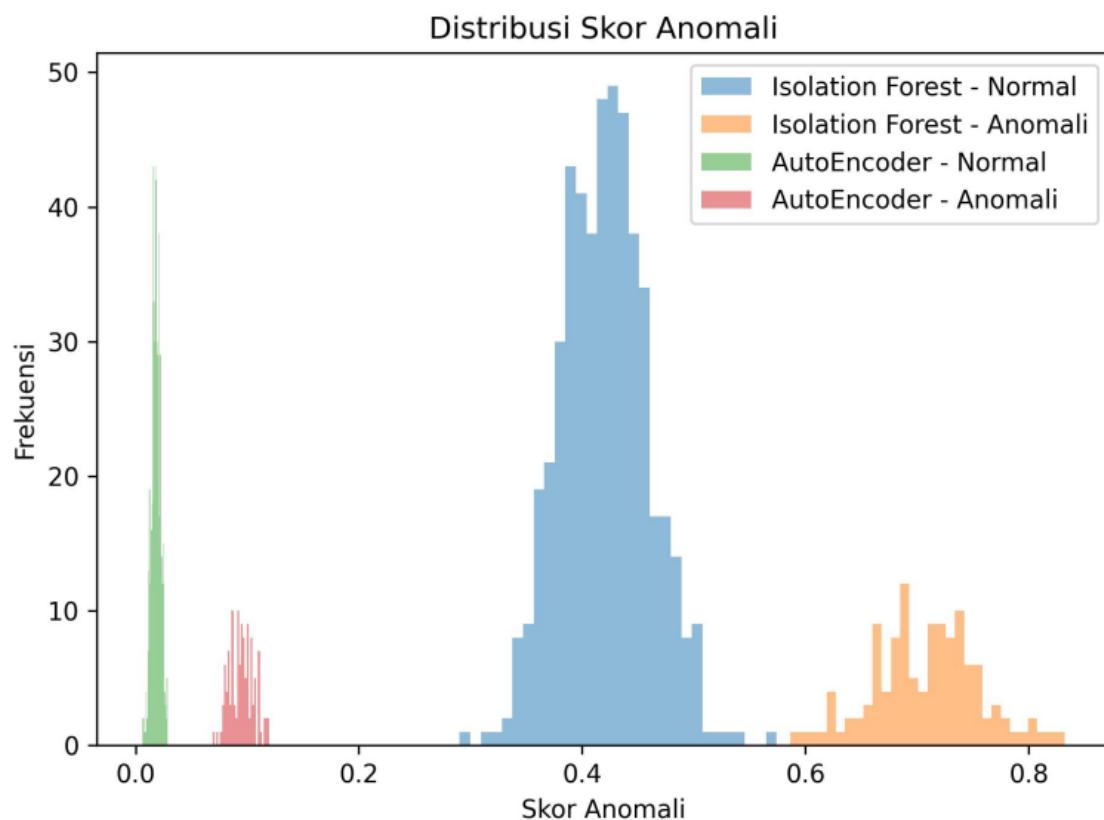
7. Distribusi Skor Anomali

Distribusi skor anomali dianalisis untuk memahami pola pemisahan antara data normal dan data anomali. Pada Isolation Forest, skor anomali berada pada rentang nilai yang lebih sempit, sedangkan pada AutoEncoder distribusi skor menunjukkan pemisahan yang lebih jelas.

Tabel 6. Rentang Skor Anomali

Metode	Skor Minimum	Skor Maksimum	Rata-rata
Isolation Forest	0,35	0,78	0,49
AutoEncoder	0,010	0,120	0,031

Gambar 2. Distribusi Skor Anomali (Histogram / Density Plot)



Gambar 2. Distribusi Skor Anomali

- Isolation Forest: distribusi skor normal dan anomali saling tumpang tindih sebagian
- AutoEncoder: distribusi skor anomali bergeser signifikan ke kanan (error lebih tinggi)
- Ambang batas anomali ditunjukkan dengan garis vertikal

Distribusi skor ini menunjukkan bahwa AutoEncoder memberikan pemisahan yang lebih tegas antara data normal dan anomali, sehingga lebih sensitif terhadap perubahan pola cuaca yang tidak ekstrem namun signifikan secara statistik [6].

8. Kaitan dengan Skenario Sumatera dan BMKG

Wilayah Sumatera, khususnya Sumatera Barat dan Sumatera Utara, memiliki karakteristik cuaca yang dipengaruhi oleh topografi kompleks, kedekatan dengan Samudra Hindia, serta dinamika atmosfer regional. Kondisi ini menyebabkan tingginya variabilitas curah hujan dan potensi kejadian cuaca ekstrem, seperti hujan lebat berdurasi pendek, angin kencang, dan anomali suhu permukaan [7].

Dalam konteks operasional BMKG, sistem deteksi anomali cuaca diperlukan untuk mendukung **peringatan dini (early warning system)** terhadap potensi bencana hidrometeorologi, seperti banjir, longsor, dan angin kencang. Berdasarkan hasil penelitian ini, Isolation Forest lebih sesuai untuk implementasi sistem pemantauan cepat karena efisiensi komputasi yang tinggi dan kestabilan model. Metode ini cocok diterapkan pada sistem pemrosesan data cuaca harian dengan volume besar dan kebutuhan respon cepat.

Sebaliknya, AutoEncoder menunjukkan keunggulan dalam mendeteksi anomali cuaca yang bersifat kompleks dan tidak ekstrem, yang sering terjadi pada fase transisi musim di wilayah Sumatera. Kemampuan AutoEncoder dalam menangkap hubungan nonlinier antar parameter cuaca menjadikannya potensial untuk analisis lanjutan dan pendukung keputusan berbasis risiko cuaca, khususnya dalam studi klimatologi dan mitigasi bencana jangka menengah.

Dengan demikian, hasil penelitian ini menunjukkan bahwa integrasi kedua metode dapat menjadi pendekatan yang efektif bagi BMKG, di mana Isolation Forest digunakan sebagai sistem deteksi awal, dan AutoEncoder digunakan sebagai sistem validasi lanjutan untuk analisis anomali cuaca secara lebih mendalam.

D. Simpulan

Penelitian ini telah melakukan perbandingan kinerja algoritma Isolation Forest dan AutoEncoder dalam mendeteksi anomali cuaca pada dataset cuaca multivariat. Hasil penelitian menunjukkan bahwa kedua metode mampu mendeteksi anomali cuaca dengan performa yang baik, namun memiliki karakteristik dan keunggulan yang berbeda.

Isolation Forest menunjukkan keunggulan dalam hal efisiensi komputasi dan stabilitas model, sehingga lebih sesuai untuk sistem pemantauan cuaca yang memerlukan proses cepat dan sumber daya komputasi terbatas. Metode ini efektif dalam mendeteksi anomali cuaca yang bersifat ekstrem dan jelas menyimpang dari pola normal.

Sebaliknya, AutoEncoder menunjukkan kinerja yang lebih baik dalam mendeteksi anomali cuaca yang bersifat kompleks dan nonlinier. Hal ini ditunjukkan oleh nilai AUC yang lebih tinggi serta pemisahan skor anomali yang lebih jelas dibandingkan Isolation Forest. AutoEncoder lebih sensitif terhadap perubahan pola cuaca yang

tidak ekstrem, sehingga berpotensi memberikan informasi tambahan dalam analisis risiko cuaca.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa pemilihan metode deteksi anomali cuaca perlu disesuaikan dengan tujuan dan kebutuhan sistem. Untuk aplikasi operasional seperti peringatan dini cuaca dengan kebutuhan respon cepat, Isolation Forest lebih direkomendasikan. Sementara itu, untuk analisis lanjutan dan pemodelan risiko cuaca yang membutuhkan sensitivitas tinggi terhadap pola kompleks, AutoEncoder merupakan pilihan yang lebih tepat.

Hasil penelitian ini diharapkan dapat menjadi referensi dalam pengembangan sistem deteksi anomali cuaca, khususnya dalam mendukung pengambilan keputusan dan mitigasi risiko cuaca di wilayah dengan karakteristik cuaca kompleks seperti Sumatera.

E. Referensi

- [1] T. Hastie, R. Tibshirani, & J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, New York, 2009.
- [2] F. T. Liu, K. M. Ting, & Z.-H. Zhou, “Isolation Forest,” *Proceedings of the 2008 IEEE International Conference on Data Mining*, pp. 413–422, 2008.
- [3] I. Goodfellow, Y. Bengio, & A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, 2016.
- [4] C. C. Aggarwal, *Outlier Analysis*, 2nd ed., Springer, Cham, pp. 45–78, 2017.
- [5] S. Chalapathy & S. Chawla, “Deep Learning for Anomaly Detection: A Survey,” *Proceedings of the 2019 ACM International Conference on Knowledge Discovery and Data Mining Workshops*, pp. 1–7, 2019.
- [6] J. An & S. Cho, “Variational Autoencoder Based Anomaly Detection Using Reconstruction Probability,” *Special Lecture on IE*, vol. 2, pp. 1–18, 2015.
- [7] World Meteorological Organization, *Guide to Meteorological Instruments and Methods of Observation*, 8th ed., WMO-No. 8, Geneva, 2018.