
Pendeteksian Anomali pada Data Operasional Migas Menggunakan Algoritma Isolation Forest dan DBSCAN

Muhammad Hafiz Shakil¹, Aditya Firman Ihsan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹hafizshak@students.telkomuniversity.ac.id,

²adityaihsan@telkomuniversity.ac.id

Abstrak

Deteksi anomali pada data operasional migas sangat penting untuk mencegah kerugian operasional, gangguan sistem, dan potensi risiko keselamatan. Penelitian ini bertujuan untuk membandingkan performa dua algoritma *unsupervised learning*, yaitu Isolation Forest dan DBSCAN, dalam mendeteksi anomali pada data historis tekanan (*pressure*) dan suhu (*temperature*) dari sektor migas. Penelitian ini menguji performa deteksi anomali menggunakan dua pendekatan, yaitu *univariate* dan *multivariate* berbasis data uji hasil *preprocessing*. Evaluasi dilakukan menggunakan metrik *silhouette score* dan *noise ratio* untuk menilai kualitas pemisahan kluster dan proporsi data yang teridentifikasi sebagai anomali. Hasil menunjukkan bahwa pendekatan *univariate*, terutama pada fitur tekanan menggunakan DBSCAN, menghasilkan *silhouette score* tertinggi sebesar 0.856403 dengan *noise ratio* yang masih wajar. Sebaliknya, pendekatan *multivariate* tidak selalu menghasilkan performa yang lebih baik dan cenderung menghasilkan *silhouette score* yang lebih rendah. Isolation Forest menunjukkan performa yang stabil di semua skenario, namun DBSCAN lebih unggul dalam memisahkan kluster secara alami pada data berdensitas tinggi. Temuan ini menunjukkan bahwa pemilihan algoritma dan pendekatan deteksi anomali sangat bergantung pada karakteristik fitur dan struktur data yang dianalisis.

Kata kunci: deteksi anomali, isolation forest, dbscan, univariat, multivariat, silhouette score

Abstract

Anomaly detection in oil and gas operational data is crucial to prevent operational losses, system disruptions, and potential safety risks. This study aims to compare the performance of two unsupervised learning algorithms Isolation Forest and DBSCAN in detecting anomalies in historical pressure and temperature data from the oil and gas sector. This research evaluates anomaly detection performance using both univariate and multivariate approaches based on preprocessed test data. The evaluation uses silhouette score and noise ratio metrics to assess clustering quality and the proportion of data identified as anomalies. The results show that the univariate approach, particularly using DBSCAN on the pressure feature, achieved the highest silhouette score of 0.856403 with a reasonable noise ratio. On the other hand, the multivariate approach did not always produce better performance and often resulted in lower silhouette scores. Isolation Forest demonstrated stable results across all scenarios, but DBSCAN outperformed in naturally separating dense clusters. These findings indicate that the choice of algorithm and detection approach highly depends on the characteristics of the features and the data structure.

Keywords: anomaly detection, isolation forest, dbscan, univariate, multivariate, silhouette score
