ABSTRACT

Anomaly detection is one of the crucial aspects in managing operational data in the oil and gas industry to prevent disruptions and system failures. This study proposes a combination of the Autoencoder and Isolation Forest algorithms to detect anomalies in oil and gas operational time series data. The Isolation Forest is applied as a pseudo-labeling method for unlabeled data, while the Autoencoder is used to learn normal patterns and identify deviations through reconstruction error.

The dataset used in this research was obtained from natural gas pipeline sensor records of a company in 2021, consisting of 62,313 data points. The study focuses on the pipeline with ***ASET_ID = 133001****, analyzing a total of 8,759 data samples with* ***TEMPERATURE*** *and* ***PRESSURE*** *as the main features.*

The evaluation was carried out using two approaches: ***Multivariate*** *and* ***Univariate****. The results show that the Univariate approach is more stable, especially when applied with the Autoencoder model. The best-performing model was obtained with the configuration of* ***Window Size = 3, dense layers 64-32, and ReLU activation function****. Performance evaluation was conducted using* ***F1-Score, Precision, and AUC*** *metrics.*

The best Autoencoder performance on the ***Temperature*** *feature achieved* ***F1-Score: 0.8177, Precision: 0.9023, Recall: 0.7476, and AUC: 0.9890****, detecting a total of 186 anomalies. On the* ***Pressure*** *feature, the model reached* ***F1-Score: 0.8432, Precision: 0.9070, Recall: 0.7879, and AUC: 0.9951****, successfully detecting 345 anomalies. These findings indicate that the combination of Isolation Forest and Autoencoder provides an effective solution for anomaly detection in oil and gas operational sensor data.*

Keywords: *Autoencoder, Isolation Forest, Anomaly Detection, Time Series Data, Oil and Gas Industry, F1-Score, ROC Curve, Pseudo-labeling, Univariate*