Abstract

In the natural gas industry, monitoring is important to reduce the risk of disaster or loss. To avoid unwanted things, a system to detect anomalies in the data generated by sensors on pipes in the industry is needed. Algorithms for anomaly detection have been widely discussed, but have not been widely applied to this industry. The goal of this research is to implement an anomaly detection system specifically designed for natural gas operational data. The multivariate time-series data used in this study was collected from multiple sensors that had placed on gas pipelines by an oil and gas company. The model is constructed using Extended Isolation Forest method. Then, a hyperparameter tuning process was employed to obtain optimal parameters for training the model. By tuning the parameter, we can ensure that the model will achieve optimal performance while having the shortest execution time. The implementation results then were evaluated and compared with the outcomes of the predecessor method, Isolation Forest. For evaluation, we used performance matrix such as recall, false alarm rate, and F1 score. The Extended Isolation Forest model achieved a recall of 0.9977, a false alarm rate of 0.0003, and an F1 score of 0.9973, slightly outperforming the standard Isolation Forest model in terms of accurate predictions, despite having a little longer execution time.

Keywords: extended isolation forest, gas pipeline operational monitoring, multivariate time-series data, anomaly detection