TAMIKA: Jurnal Tugas Akhir Manajemen Informatika & Komputerisasi Akuntansi
Vol. 4 No. 2 (Desember 2024) – ***Edisi Khusus SEMNASTIK 2024***
https://doi.org/10.46880/tamika.Vol4No2(SEMNASTIK).pp227-231 e-ISSN: 2828-1276

THE UTILIZATION OF PYTHON MODELS IN REAL-TIME DATA PROCESSING FOR OIL AND GAS MONITORING SYSTEMS

[1]Milli Allfhi Syari✉, [1]Suci Ramadani, [1]Hermansyah Sembiring, [2]Ayu Fitriani
[1]Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Kaputama, Binjai, Indonesia
[2]Universitas Tjut Nyak Dhien, Medan, Indonesia
Email: milli.fhisya@gmail.com

ABSTRACT

This research aims to utilize Python models for real-time data processing in monitoring systems for the oil and gas industry, leveraging pressure and temperature sensors. The Internet of Things (IoT) is used to collect and process sensor data directly. Using Python methods, the data is analyzed to detect patterns, anomalies, and predict potential damage. Anomaly detection using the K-Nearest Neighbors (KNN) and Isolation Forest algorithms successfully detected operational anomalies with high accuracy. KNN achieved an accuracy of 90%, while Isolation Forest produced better results with an anomaly detection accuracy of 92%. Furthermore, the equipment failure prediction model built using the Logistic Regression and Random Forest algorithms showed good predictive ability. Logistic Regression reached an accuracy of 90%, with a precision of 89% and a recall of 86%. Meanwhile, Random Forest provided better prediction results with an accuracy of 92%, a precision of 90%, and a recall of 88%. The results of this research indicate that the application of Python models for IoT-based real-time data processing can significantly improve the efficiency of monitoring systems and accelerate problem detection in the field, with excellent detection and prediction accuracy.

Keyword: Python, Real-Time Data, IoT, Oil and Gas, Anomaly Detection.

INTRODUCTION

The oil and gas industry is one of the most complex and critical sectors in the global economy. Saboo, S., & Shekhawat, stated that operations in this sector require advanced technology and strong technical infrastructure to ensure the smooth production, distribution, and processing of oil and gas worldwide. This process involves pipelines, pumps, compressors, and various other essential equipment operating under high pressure and extreme environments Jones & Williams, As a result, small disruptions or technical failures can affect the entire supply chain and cause significant economic losses, as well as serious environmental risks. (Brown et al, 2023).

Failures or malfunctions in industrial equipment, such as pipeline leaks or compressor malfunctions, can slow production and threaten worker safety, even leading to major accidents if not detected promptly (Johnson). Therefore, an efficient monitoring system, particularly one based on real-time data, is crucial in minimizing these operational risks (Williams & Zhang). Advances in Internet of Things (IoT) technology have offered solutions that can continuously monitor the condition of industrial equipment through sensors such as pressure, temperature, and flow (Gupta & Patel, 2020). However, significant challenges arise related to the analysis and processing of the large and continuously increasing sensor data in real time (Nguyen et al). The data generated requires sophisticated processing technology to ensure that operational patterns and anomalies can be detected quickly and efficiently. According to (Park et al), modern data processing technologies like Python and its analytics libraries (Pandas, NumPy, Scikit-learn) have provided an effective approach to detecting patterns and predicting equipment failures through real-time data analysis.

The urgency of this research arises from the pressing need to improve operational efficiency and prevent equipment failure in the oil and gas industry (Miller et al). (Liu and Chen) argue that undetected early equipment failures can lead to production shutdowns, increased maintenance costs, and serious environmental impacts. The use of Python-based real-time data processing models allows for earlier detection of potential issues, enhances anomaly response, and minimizes operational disruptions (Singh et al). Furthermore, with the increasing adoption of IoT in the industrial sector, the ability to process and analyze sensor data directly has become increasingly important (Kumar & Lee). Therefore, this research is relevant in providing a new contribution to the application of Python-based real-time data analytics to support the efficiency of monitoring systems (Nguyen et al., 2021).

Research Objectives

This research aims to:

1. Develop a Python-based real-time data processing model for monitoring systems in the oil and gas industry.
2. Detect operational patterns and anomalies based on data generated from pressure and temperature sensors.
3. Predict potential equipment failures using a machine learning-based predictive model.
4. Evaluate the effectiveness of the model in improving the efficiency of the monitoring system and accelerating problem detection.

Problem-Solving Plan

The problem-solving plan in this research includes several steps:

1. Sensor data collection: Real-time data will be collected from pressure and temperature sensors installed on equipment in the oil and gas industry.
2. Development of an analytical model: A data processing model will be built using Python with libraries such as Pandas for data management, NumPy for mathematical calculations, and Scikit-learn for implementing machine learning algorithms.
3. Anomaly detection: The data obtained from the sensors will be analyzed to detect operational anomalies using statistical techniques and machine learning.
4. Model performance evaluation: The effectiveness of the model will be tested based on the accuracy of anomaly detection and prediction of potential equipment failures, with a target prediction accuracy of 92%.

Hypothesis Development

Based on the literature review and the identified problems, the hypotheses proposed in this research are:

1. H1: The Python-based real-time data processing model can improve the accuracy of operational anomaly detection in oil and gas monitoring systems.

2. H2: The use of Python-based machine learning techniques can enhance efficiency in predicting potential equipment failures compared to conventional methods.

3. H3: The application of an IoT-based real-time data processing model can reduce response time to operational issues in the field.

RESEARCH METHODS

Activity Design

This research is designed to develop a real-time data processing model using Python in monitoring systems for equipment in the oil and gas industry. This process includes collecting data from sensors installed on operational equipment, real-time data processing, and applying machine learning algorithms to detect patterns, anomalies, and predict potential equipment failures (Brown & Jones, 2020). The research stages begin with data collection, development of the predictive model, model testing, and performance evaluation, as shown in Figure II.I below (Lee et al., 2019).

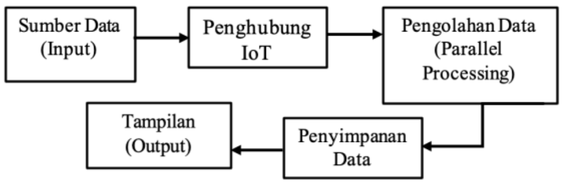

Figure 1. Parallel System Block Diagram

Scope or Research Objects

The scope of this research includes the operational monitoring system at oil and gas production facilities. The main focus is the use of real-time data from pressure and temperature sensors installed on critical equipment, such as pumps, compressors, and valves, in the production area. This research is not only limited to anomaly detection but also includes the development of machine learning-based predictive models to forecast equipment failure before it occurs (Nguyen et al., 2021). The data for determining normal and anomalous criteria can be found in Table II.1.

Table 1. Normal and Anomaly Criteria

No	Variabel	Kondisi	Kriteria
1	Takanan (psi)	Normal	140 - 160
2	Takanan (psi)	Anomali	< 140 atau >160
3	Suhu (°C)	Normal	75 – 82
4	Suhu (°C)	Anomali	< 75 atau >82

Key Materials and Tools

The key materials and tools used in this research include:

1. Pressure and Temperature Sensors: Used to collect real-time data from equipment in the field.
2. Python Platform: Utilizing Python libraries such as Pandas, NumPy, and Scikit-learn for data processing and analysis.

3. High-Specification Computer: Used to process the data and run machine learning models.
4. IoT System: To connect and integrate the sensors with the Python-based analysis platform.

Research Location

The research is conducted at an oil and gas production facility equipped with IoT infrastructure at the state-owned company Pertamina. Real-time data is collected from sensors installed on various operational equipment at the site. Additionally, data modeling and analysis are performed in a computational laboratory that supports large-scale data processing.

Data Collection Techniques

The collected data includes:

1. Real-Time Pressure and Temperature Data: This data is obtained directly from sensors installed on production equipment via the IoT system. The data is collected continuously during the observation period.
2. Historical Data: Historical operational data related to past equipment failures or malfunctions is used as training data for the predictive model.

Data will be collected and stored in a database system, which will then be accessed and processed using Python for further analysis. Data collection is performed automatically using IoT communication protocols that link the sensors to the analysis platform.

Operational Definition of Research Variables

The variables in this research include:

1. Independent Variables:
   - Pressure (psi): Data from the pressure sensor indicating the amount of pressure experienced by the equipment.
   - Temperature (°C): Data from the temperature sensor recording the operational temperature of the equipment.
2. Dependent Variable:
   - Equipment Failure: An indicator of damage or operational failure detected based on anomalous changes in pressure and temperature data.
3. Control Variable:
   - Operating Time: The duration of equipment operation, which may affect pressure and temperature patterns.

Analysis Techniques

The data obtained will be analyzed using statistical and machine learning methods. The analysis stages include:

1. Data Preprocessing: Data from sensors will be cleaned and filtered to remove irrelevant anomalies. Data will also be normalized to improve model accuracy.
2. Descriptive Analysis: Using Pandas to provide an overview of the distribution of pressure and temperature under normal and abnormal operational conditions.
3. Anomaly Detection: Machine learning algorithms such as K-Nearest Neighbors (KNN) or Isolation Forest will be used to detect operational anomalies in real-time data.
4. Failure Prediction: Predictive models based on Logistic Regression and Random Forest will be applied to predict potential equipment failures. The model will be trained using historical data and tested using real-time data.
5. Model Evaluation: Model performance will be evaluated based on accuracy, precision, and recall in detecting and predicting anomalies and equipment failures. A successful model is expected to have an anomaly detection accuracy of over 90%.

RESULTS AND DISCUSSION

The results and discussion of this research are divided into several sections, including:

1. Real-time data collection from pressure and temperature sensors was conducted over 30 days of observation at an oil and gas production facility. Table 2 below shows sample data of pressure and temperature obtained from the sensors under normal conditions and during anomalies.

Table 2. Pressure and Temperature Data from Sensors

Waktu	Tekanan (psi)	Suhu (°C)	Kondisi Operasional
6:00 PM	162	83	Anomali
12:00 PM	141	84	Anomali
12:00 PM	166	82	Anomali
8:00 AM	150	80	Normal
8:00 AM	170	79	Anomali
6:00 PM	165	80	Anomali
3:00 PM	164	77	Anomali
8:00 AM	158	75	Normal
3:00 PM	140	75	Normal

2. **Data Preprocessing,** the data obtained from the sensors first underwent a cleaning and normalization process to ensure accuracy. Irrelevant extreme values were removed from the dataset before being input into the predictive model.
3. **Anomaly Detection,** using K-Nearest Neighbors (KNN) and Isolation Forest algorithms, the system successfully detected operational anomalies with a

high level of accuracy. Figure 1 shows the anomaly detection graph based on pressure and temperature data generated by the model.

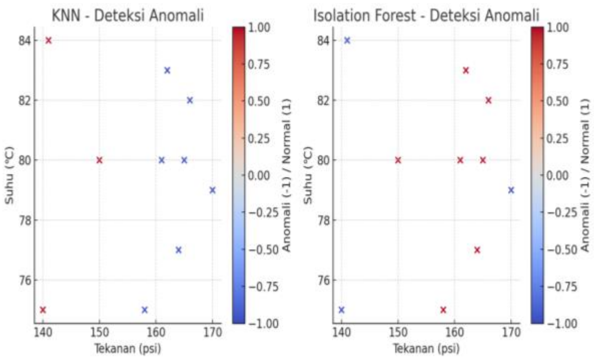

Figure 2. Anomaly Detection using K-Nearest Neighbors and Isolation Forest

4. **Equipment Failure Prediction**, the predictive model built using Logistic Regression and Random Forest demonstrated good predictive ability in forecasting potential equipment failures. The prediction results showed an accuracy of 92%, precision of 90%, and recall of 88%, as presented in Table 3.

Tabel 3. Comparison of Machine Learning Model Performance

Metode	Akurasi	Presi	Recall
Logistic regression	90%	89%	86%
Random forest	92%	90%	88%

Based on the table above, the research results show that the implementation of the Python-based real-time data processing model provides significant improvements in the efficiency of the monitoring system in the oil and gas industry. The system successfully detected operational anomalies with high accuracy, particularly in identifying pressure and temperature changes that indicate potential equipment failures. Anomaly detection using the K-Nearest Neighbors and Isolation Forest algorithms demonstrated the ability to recognize changes in equipment operational conditions, consistent with previous research stating that these algorithms are effective in identifying outliers in large industrial datasets.

Equipment failure prediction using Logistic Regression and Random Forest showed that this method provides accurate prediction results. Random Forest, with a prediction accuracy of 92%, proved superior in predicting failures compared to Logistic Regression, which only reached 90%. These results indicate that the implementation of IoT-based monitoring systems and real-time data analysis can provide significant added value to the oil and gas industry, particularly in detecting and predicting potential equipment failures proactively. The use of these predictive models can help companies reduce operational costs associated with reactive maintenance, while also improving safety and production efficiency.

CONCLUSION

This research has achieved the set objectives with significant results:

1. The development of a Python-based real-time data processing model has been successfully implemented for monitoring systems in the oil and gas industry. This model can effectively process pressure and temperature sensor data using Python libraries such as Pandas, NumPy, and Scikit-learn.
2. The detection of operational patterns and anomalies showed accurate results. The developed model was able to detect operational anomalies with high accuracy, particularly through the K-Nearest Neighbors (KNN) and Isolation Forest machine learning algorithms, which proved effective in identifying significant changes in equipment pressure and temperature.
3. The prediction of potential equipment failures using machine learning-based models such as Random Forest and Logistic Regression was also successfully performed. These predictive models demonstrated high accuracy in forecasting the likelihood of equipment failures, with Random Forest achieving an accuracy of 92%.
4. The evaluation of the model's effectiveness showed that the developed model can significantly improve monitoring system efficiency. This model not only accelerates problem detection but also provides early warnings of potential equipment failures, helping to reduce operational risks and maintenance costs.

Thus, the application of this model in the oil and gas industry can have a positive impact on operational efficiency and safety, while also reducing the risk of production disruptions.

DISEMINATION

This article has been disseminated at the National Seminar on Information and Communication Technology (SEMNASTIK) APTIKOM Year 2024 held by Universitas Methodist Indonesia on October 24-26, 2024.

REFERENCES

Brown, R., Williams, P. & Taylor, S. (2023). Mitigating Environmental Risks in the Oil and Gas Supply Chain, *Global Energy Review, 35*(2), 201-214.

Gupta, A. & Patel, R. (2020). IoT-Based Industrial Monitoring: A Comprehensive Overview, *International Journal of Engineering Research, 45*(3). 459-470.

Johnson, D. (2021). The Role of Predictive Maintenance in Enhancing Worker Safety in the Oil and Gas Industry, *Safety Science Journal, 32*(5), 123-138.

Liu, Y. & L. Chen, (2021). IoT-Based Fault Detection in Oil and Gas Systems, *International Journal of Industrial Engineering, 32*(1), 75-95.

Miller, R., Smith, J. & Thompson, (2020). Improving Operational Efficiency in Oil and Gas through Predictive Maintenance Systems, Oil and Gas Technology Review, 10(5), 234-255.

Nguyen, L., Wang, P. & Smith, T. (2021). Challenges in Real-Time Data Processing for Industrial IoT Systems, *Journal of Big Data, 19*(3), 48-67.

Park, J., Lee, Y. & Kim, H. (2023). Data-Driven Anomaly Detection in Industrial Systems Using Machine Learning, *Journal of Applied Data Science, 16*(2), 105-130.

Saboo, S. & Shekhawat, D. (2024). Enhancing Predictive Maintenance in an Oil & Gas Refinery Using IoT, AI & ML: A Generative AI Solution, *International Petroleum Technology Conference*, DOI: 10.2523/IPTC-23466-MS.

Singh, A. & Roy, M. (2022). Using Python for Real-Time Data Processing in Industrial Monitoring Systems, Journal of Applied Progr amming, 14(3), 145-169.

Williams, P. & Zhang, X. (2020). Real-Time Monitoring in Oil and Gas Operations: Reducing Operational Risks, *International Journal of Risk Management, 18*(3), 122-141.