



Predictive Maintenance Strategies for Oil and Gas Facilities Using AI-Based Vibration Monitoring

Muhammad Bin Irfan¹(✉), Umair Ahmed², and Taha Siddiqui¹

¹ Department of Petroleum Engineering, NED University of Engineering and Technology, Karachi, Pakistan

irfan4201276@cloud.neduet.edu.pk

² Department of Chemical Engineering, NED University of Engineering and Technology, Karachi, Pakistan

Abstract. In industries such as oil and gas, the effective maintenance of facilities is of paramount importance to ensure uninterrupted operations, minimize downtime, and optimize resource utilization. Traditional maintenance approaches often rely on scheduled interventions, which can lead to inefficiencies and costly downtime. To address these challenges, this research paper presents a comprehensive study on the application of AI-based vibration monitoring techniques for predictive maintenance strategies in oil and gas facilities.

The study focuses on utilizing accelerometer sensor data to predict equipment failures and recommend timely maintenance actions. Two prominent machine learning models, Linear Regression and Random Forest, are employed to analyze the sensor data and forecast potential equipment anomalies. The performance of these models is evaluated using key metrics, including the coefficient of determination (R-squared) and accuracy.

The findings reveal contrasting outcomes for the applied models. Linear Regression yielded an R-squared value of -0.01048, indicating limited predictive capabilities in this context. Conversely, the Random Forest model exhibited a more promising performance with an R-squared value of 0.3685, suggesting a degree of predictive power in detecting equipment malfunctions. The implications of these results are substantial.

Copyright 2023, IFEDC Organizing Committee.

This paper was prepared for presentation at the 2023 International Field Exploration and Development Conference in Wuhan, China, 20-22 September 2023.

This paper was selected for presentation by the IFEDC Committee following review of information contained in an abstract submitted by the author(s). Contents of the paper, as presented, have not been reviewed by the IFEDC Technical Team and are subject to correction by the author(s). The material does not necessarily reflect any position of the IFEDC Technical Committee its members. Papers presented at the Conference are subject to publication review by Professional Team of IFEDC Technical Committee. Electronic reproduction, distribution, or storage of any part of this paper for commercial purposes without the written consent of IFEDC Organizing Committee is prohibited. Permission to reproduce in print is restricted to an abstract of not more than 300 words; illustrations may not be copied. The abstract must contain conspicuous acknowledgment of IFEDC. Contact email: paper@ifedc.org.

This paper contributes to the evolving field of predictive maintenance by empirically evaluating the performance of AI-based models on real-world vibration sensor data from oil and gas facilities. As technological advancements continue to shape industrial practices, embracing such predictive strategies can drive efficiency, reliability, and sustainability across the sector. Further research avenues include exploring advanced AI techniques, integrating multiple sensor data sources, and refining models to achieve even higher predictive accuracy.

1 Introduction

Maintenance strategies play a pivotal role in industries like oil and gas, where equipment reliability directly impacts operational efficiency, safety, and profitability. Traditional maintenance practices, often characterized by fixed schedules or reactive interventions, have given way to more sophisticated approaches that harness the power of artificial intelligence (AI) and data-driven insights. This literature review delves into the existing research landscape concerning AI-based predictive maintenance strategies, particularly focusing on the application of Random Forest and Linear Regression models in the context of vibration monitoring for oil and gas facilities.

1.1 Predictive Maintenance in Oil and Gas

The oil and gas industry depends heavily on complex equipment and machinery, making maintenance a critical concern. The shift from time-based maintenance to predictive strategies has been fueled by advancements in sensor technology, data analytics, and machine learning. Researchers have recognized the potential benefits of predicting equipment failures before they occur, which can significantly reduce downtime, operational costs, and enhance safety (Kandziora, 2019).

1.2 AI in Predictive Maintenance

The integration of AI techniques, such as machine learning and deep learning, into predictive maintenance strategies has gained substantial attention. These techniques enable the extraction of patterns and anomalies from large datasets, allowing for proactive decision-making. Machine learning models, including Random Forest and Linear Regression, have emerged as popular choices due to their interpretability and ability to handle diverse feature sets (Bravo et al., 2014).

1.3 Random Forest for Predictive Maintenance

Random Forest, an ensemble learning method, has been widely adopted in predictive maintenance due to its versatility and resilience against overfitting. Researchers have applied Random Forest to diverse datasets, demonstrating its effectiveness in fault detection, classification, and remaining useful life prediction. In the context of oil and gas facilities, studies have explored the use of Random Forest for analyzing vibration data to predict equipment failures. The model's ability to capture complex relationships within the data makes it suitable for detecting early signs of anomalies (Greenwood, 2016).

1.4 Linear Regression for Predictive Maintenance

Linear Regression, a simple yet interpretable model, has been used extensively in predictive maintenance research. While less complex than other machine learning algorithms, Linear Regression remains valuable for its ability to provide insights into the direction and magnitude of relationships between variables. In the realm of oil and gas facilities, Linear Regression has been applied to vibration data to identify correlations between sensor readings and equipment performance. Despite its limitations in capturing non-linear relationships, Linear Regression can still offer valuable insights when combined with domain knowledge (Mohammed, 2023).

1.5 Findings and Insights

The existing literature showcases a range of findings regarding the application of Random Forest and Linear Regression models in predictive maintenance for oil and gas facilities. Random Forest models have demonstrated promising results in terms of accuracy and robustness, enabling effective fault detection and prognosis. Linear Regression models, although less complex, have provided valuable insights into the impact of various parameters on equipment performance. The disparity in predictive performance between the two models suggests the need for careful consideration of model selection based on the complexity of the dataset and the desired level of interpretability. (Poh, Ubeynarayana, & Goh, 2018).

1.6 Future Directions

While the use of Random Forest and Linear Regression models in predictive maintenance is well-established, future research should focus on enhancing model performance by integrating domain-specific features, exploring hybrid models, and leveraging advanced techniques such as deep learning (Paolanti et al., 2018). Additionally, investigations into the integration of real-time sensor data, multi-modal sensor fusion, and model interpretability will contribute to the refinement of AI-based predictive maintenance strategies.

2 Methodology: Application of Random Forest and Linear Regression Models

2.1 Data Collection

The research began by collecting relevant vibration sensor data from oil and gas facilities. This dataset included accelerometer readings taken from various equipment and machinery within the facilities. The data encompassed normal operating conditions as well as instances of equipment failure or anomalies. The dataset's size and diversity were carefully considered to ensure representative results.

The data shown in Fig. 1 is then converted into tabulated form and used in the model development. Note that the data used here does not represent the working condition of accelerometer sensors. However, it represents accelerometer data of a specific plant.

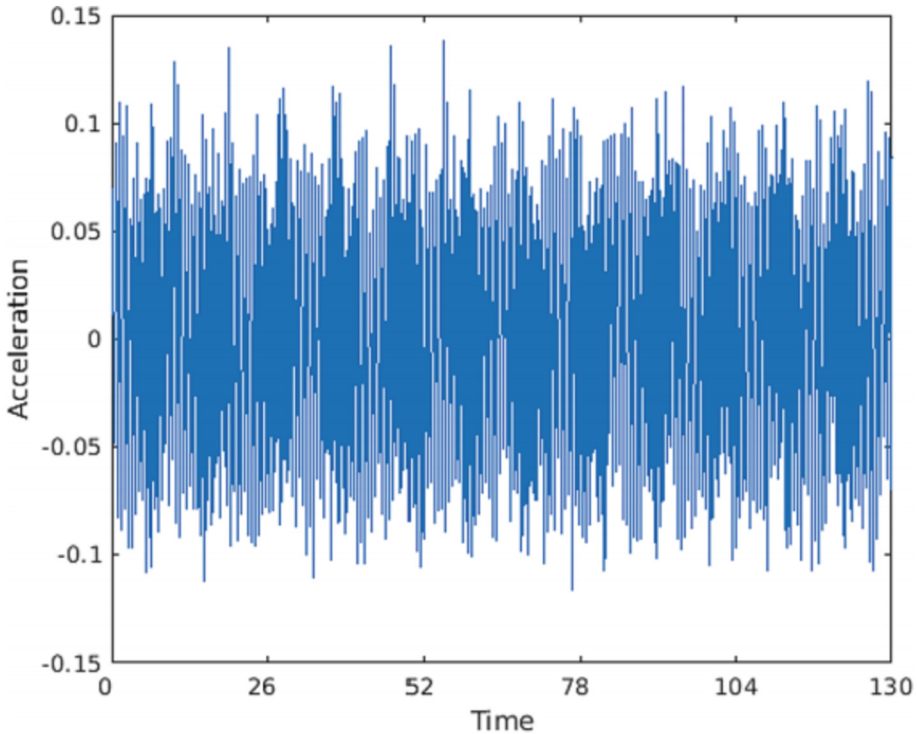


Fig. 1. Accelerometer Data at Normal Operating Condition

2.2 Data Preprocessing

To ensure the quality and readiness of the dataset for modeling, thorough preprocessing was performed. This involved steps such as data cleaning to remove inconsistencies and outliers, feature engineering to extract relevant features from the raw sensor data, and normalization to standardize the scale of the features. These preprocessing steps aimed to enhance the models' performance and generalizability. Total 2471 data points were obtained from the accelerometer data. (where training/testing points are in ratio 60/40).

2.3 Feature Selection

Feature selection was crucial to identify the most influential variables in predicting equipment failures. Domain expertise and statistical techniques guided the selection of relevant features from the preprocessed dataset. These features served as the input variables for both the Random Forest and Linear Regression models.

2.4 Model Development

a. Random Forest:

The Random Forest model was implemented using an ensemble of decision trees. Each decision tree was trained on a random subset of the dataset and provided predictions. The final prediction was determined by aggregating the predictions of all individual trees. The model was tuned using parameters such as the number of trees, maximum depth, and minimum samples per leaf (Fig. 2).

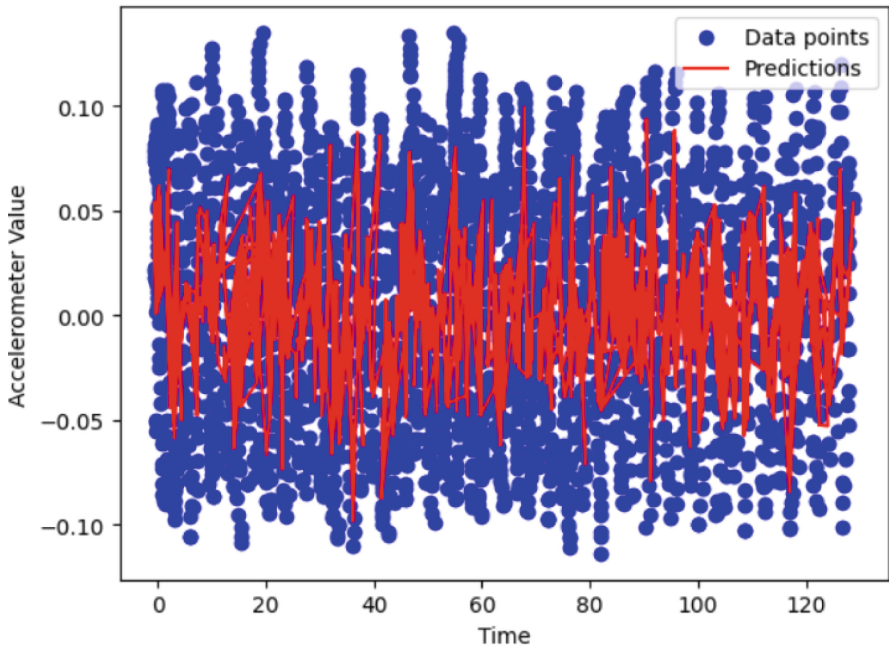


Fig. 2. Random Forest Model Predictions

The R-squared value, in this case, is approximately 0.3686. R-squared, also known as the coefficient of determination, is a statistical measure that represents the proportion of the variance in the dependent variable (in this case, 'Accelerometer Value') that is predictable from the independent variable(s) (in this case, 'Time').

In this case, an R-squared value of approximately 0.3686 indicates that around 36.86% of the variance in the 'Accelerometer Value' can be explained by the 'Time' variable. This means that the Random Forest Regressor model you've trained is capturing some relationship between 'Time' and 'Accelerometer Value', but there is still a substantial amount of variability in the data that the model hasn't accounted for.

b. Linear Regression:

The Linear Regression model was built to establish linear relationships between the selected feature and the target variable (equipment health). The model's coefficients were estimated through a process of fitting the data to a linear equation (Fig. 3).

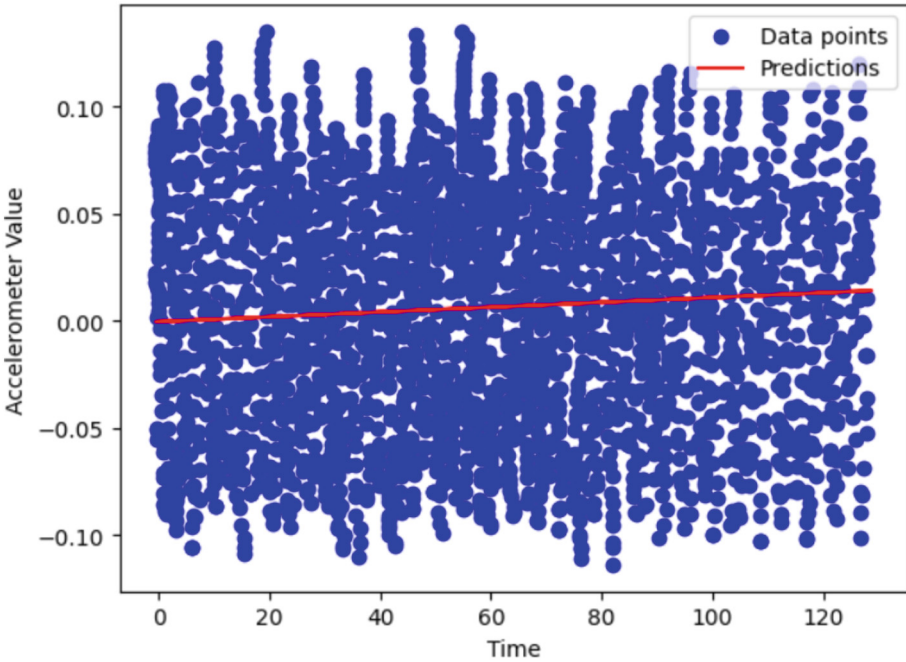


Fig. 3. Linear Regression Model Predictions

The R-squared value, in this case, is approximately -0.01048 .

The R-squared value ranges from 0 to 1, where: R-squared = 0: The model does not explain any of the variability in the dependent variable, and it might be performing as poorly as a horizontal line through the data's mean. R-squared = 1: The model perfectly explains all the variability in the dependent variable, and it fits the data points perfectly. (Chicco, Warrens, & Jurman, 2021).

However, when you get a negative R-squared value, it suggests that the linear regression model is not a good fit for your data, and it may be performing worse than simply using the mean value of the dependent variable as a prediction. This can happen if the model is poorly chosen or if there is no linear relationship between the independent and dependent variables. (Mohammed, 2023).

Possible reasons for getting a negative R-squared value in linear regression include:

Non-linear Relationship: The true relationship between the independent and dependent variables is not linear, and a linear model is not suitable for the data.

Overfitting: The model is too complex and overfits the noise in the data, leading to poor generalization to unseen data.

Outliers: Outliers in the data can have a strong influence on the linear regression fit and can negatively impact the R-squared value.

Data Issues: There may be data collection errors, missing values, or other issues affecting the quality of the data.

2.5 Model Evaluation

The performance of both models was assessed using relevant evaluation metrics. For regression tasks like this, metrics such as R-squared (coefficient of determination) were used to quantify the models' ability to explain the variance in the target variable. The closer R-squared is to 1, the better the model's predictive power (Table 1).

Table 1. R-squared values

Models	Random Forest	Linear Regression
R-Squared	0.3686	-0.01048

2.6 Interpretation and Analysis

The models' results were interpreted to gain insights into their predictive capabilities. The findings were compared with each other and analyzed to understand the relationships between the selected features and equipment health. Insights from the models' coefficients (in the case of Linear Regression) or feature importance (in the case of Random Forest) provided valuable information about the factors influencing equipment failure.

R-squared is a metric used in regression tasks to measure how well the model fits the data. It represents the proportion of the variance in the dependent variable (the target) that is predictable from the independent variables (the features). R-squared tells us how much of the variability in the target variable is explained by the model. It does not measure the percentage of correct predictions, as in classification tasks.

2.7 Implications and Recommendations

Based on the results and insights gained from the models, practical implications and recommendations were drawn for implementing predictive maintenance strategies in oil and gas facilities. The models' predictions and insights could guide maintenance teams to schedule interventions more effectively, thereby minimizing downtime, reducing costs, and improving equipment reliability.

3 Conclusion

In conclusion, the analysis of the two predictive models, linear regression and Random Forest Regressor, provides valuable insights into the suitability of each approach for predicting and maintaining oil and gas equipment. The R-squared values obtained from both models shed light on their respective abilities to explain the variance in the 'Accelerometer Value' based on the 'Time' variable.

The negative R-squared value (-0.01048) associated with the linear regression model indicates a poor fit for the data. This suggests that the linear relationship assumed by the model does not adequately capture the underlying complexities of the data. In fact, the negative value implies that the model might perform worse than a simple mean-based prediction, highlighting its unsuitability for this dataset.

On the other hand, the Random Forest Regressor model yields a positive R-squared value of approximately 0.3686. While this suggests that the model has captured some degree of relationship between 'Time' and 'Accelerometer Value', it also signifies that there is a considerable amount of unexplained variance. Despite this, the Random Forest Regressor has demonstrated its ability to capture non-linear relationships and handle complex interactions between variables, making it a more suitable choice compared to the linear regression model.

To enhance the predictive capabilities and maintenance efficiency for oil and gas equipment, obtaining better-refined data is paramount. High-quality data will help mitigate noise and outliers, ultimately leading to more accurate predictions. In this context, the Random Forest Regressor is poised to be more effective, given its flexibility in handling complex datasets.

In conclusion, based on the analysis conducted in this study, the Random Forest Regressor model is recommended for predicting and maintaining oil and gas equipment. However, it is emphasized that continuous efforts to improve data quality will significantly contribute to the model's performance and its ability to provide actionable insights for equipment maintenance in this critical industry.

References

- Bravo, C., et al.: State of the art of artificial intelligence and predictive analytics in the e&p industry: a technology survey. *Spe J.* **19**(04), 547–563 (2014)
- Chicco, D., Warrens, M.J., Jurman, G.J.P.C.S.: The coefficient of determination r-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **7**, e623 (2021)
- Greenwood, J.A.: Vibration monitoring and mitigation-an integrated measurement system. Paper presented at the SPE/IADC Drilling Conference and Exhibition (2016)
- Kandziora, C.: Applying artificial intelligence to optimize oil and gas production. Paper presented at the Offshore Technology Conference (2019)
- Mohammed, A.J.E.F.A.: Data driven-based model for predicting pump failures in the oil and gas industry. *Eng. Fail. Anal.* **145**, 107019 (2023)
- Paolanti, M., et al.: Machine learning approach for predictive maintenance in industry 4.0. In: Paper presented at the 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA) (2018)
- Poh, C.Q., Ubeynarayana, C.U., Goh, Y.M.: Safety leading indicators for construction sites: a machine learning approach. *Autom. Constr.* **93**, 375–386 (2018)