Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

Z-Score Normalized Machine Learning Approach for Predictive Maintenance Optimization in Gas Treatment Plants

Femi Adeoye Alabi[1]
Department of Electrical/Electronic Engineering, University of Port Harcourt.
falabious@yahoo.com

Bamidele Stephen Omoyajowo[2] (Ph.D)
Department of Science Education, Faculty of Education Lead City University
omoyajowo.bamidele@lcu.edu.ng
https://orcid.org/0009-0009-7824-2580

Nkolika O. Nwazor (Ph.D)[3]
Department of Electrical/Electronic Engineering, University of Port Harcourt.
nkolika.nwazor@uniport.edu.ng

Abstract

The increasing global demand for cleaner energy has positioned natural gas as a vital resource, requiring gas treatment plants (GTPs) to operate efficiently and reliably. However, the harsh operational conditions of GTPs—characterized by high pressures, temperatures, and corrosive gases—expose critical equipment to frequent wear and tear, leading to unexpected failures and costly downtimes. Traditional maintenance strategies such as reactive and preventive maintenance have proven inadequate in addressing these challenges due to their reliance on failure occurrence and fixed service intervals, respectively. This study aimed to evaluate the performance of selected machine learning (ML) models—Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)—for predictive maintenance using operational datasets processed through z-score normalization. Historical metering data obtained from Total Energies EP Nigeria Limited (2019–2024) was preprocessed to remove outliers, engineered to create predictive features, and scaled for uniformity. The models were assessed based on accuracy, precision, recall, and F1-score. Results revealed that all four models performed optimally after z-score normalization, with SVM achieving the highest and most consistent performance across all metrics. KNN also demonstrated robust results, particularly in adapting to different feature scaling methods. Based on these findings, the study concludes that SVM is the most effective model for predicting maintenance needs in GTPs. It is recommended that gas treatment facilities adopt SVM-based predictive maintenance systems to enhance reliability, reduce operational costs, and improve equipment lifespan.

Keywords: Predictive maintenance, gas treatment plants, machine learning, z-score normalization, Support Vector Machine.

Introduction

Energy continues to be the driving force behind economic development, technological innovation, and improved living standards. Over the past century, energy demand has risen sharply, reaching 552 quadrillion British thermal units (quads) globally in 2016—a trend propelled by rapid urbanization, population growth, and growing reliance on technology (Pandey et al., 2020). In that same year, the oil and gas (O&G) industry was responsible for supplying 55% of the global energy demand, a figure projected to increase to 57% by 2040 (Ediger et al., 2023).

Among fossil fuels, natural gas is gaining attention as a preferred energy source due to its lower carbon emissions, high efficiency, and versatility across power generation, transportation, and industrial applications (Gao et al., 2022). It's relatively cleaner combustion—producing nearly 50% less carbon dioxide than coal—makes it an environmentally sustainable option for electricity generation (Mohammad et al., 2021). However, delivering natural gas in a usable form requires a well-integrated infrastructure that includes extraction, transport, and crucially, purification at gas treatment plants (GTPs). These facilities are tasked with removing harmful compounds such as hydrogen sulfide (H_2S), carbon dioxide (CO_2), and water vapor to ensure the gas meets safety and quality standards (Mokhatab et al., 2018; Wilson et al., 2023).

The efficiency of GTPs heavily depends on the reliability of core equipment—compressors, scrubbers, separators, and heat exchangers—which routinely endure harsh conditions including high pressure, elevated temperatures, and corrosive substances (Poe & Mokhatab, 2017). Failures in any of these components can trigger major safety incidents, environmental hazards, and costly production halts. For instance, hydrogen sulfide has been shown to significantly shorten equipment lifespan, thereby escalating maintenance costs (Al-Janabi, 2020).

To sustain operational reliability, GTPs have historically employed reactive maintenance (RM) and preventive maintenance (PM). RM addresses breakdowns after they occur, often resulting in excessive downtime and repair expenses (Ucar et al., 2024; Achouch et al., 2022). PM, which relies on scheduled checks, aims to prevent failure through routine servicing, but it can be inefficient when servicing is done before it's truly needed, leading to resource wastage (Yang et al., 2021). Moreover, traditional maintenance strategies are estimated to account for over one-third of plant operating costs (Francesco et al., 2020).

As such, the industry is transitioning towards Predictive Maintenance (PdM)—a forward-looking strategy that leverages real-time data, sensor feedback, and historical patterns to forecast equipment health and prevent failure (Ucar et al., 2024). With the rise of Industry 4.0 and the Internet of Things (IoT), PdM has become more feasible, as intelligent systems driven by artificial intelligence (AI) and machine learning (ML) can now process large datasets to identify anomalies and degradation trends (Silvestri et al., 2020; Arena et al., 2024).

ML techniques, including algorithms like Support Vector Machines (SVM), Random Forest (RF), and Artificial Neural Networks (ANN), have been applied successfully in a wide range of engineering applications, particularly in failure prediction and classification tasks (Sircar et al., 2021; Aliyu et al., 2022). However, one critical challenge that influences ML model accuracy is the presence of outliers and differences in feature scales. These inconsistencies can distort the training process, making models unreliable. Hence, effective preprocessing, particularly through techniques like z-score normalization, is essential for standardizing feature distributions and enhancing predictive performance.

This study aims to evaluate the performance of selected machine learning models for predictive maintenance in gas treatment plants using operational data influenced by outliers and normalized through z-score feature scaling.

Statement of the Study

Gas treatment plants are critical to the natural gas supply chain, ensuring that raw gas is refined to meet safety and quality standards for end users. However, the equipment within these facilities—such as compressors, heat exchangers, separators, and scrubbers—is routinely exposed to intense operating conditions including high pressure, extreme temperatures, and corrosive gases (Poe & Mokhatab, 2017). These harsh conditions accelerate wear and tear, making the reliability of plant operations heavily dependent on effective maintenance strategies.

Traditionally, gas treatment facilities have relied on reactive and preventive maintenance methods. Reactive maintenance, which addresses faults only after failure has occurred, often leads to unexpected downtimes and high repair costs (Abidi et al., 2022). Preventive maintenance, while scheduled and proactive, is typically based on historical trends and fixed intervals. This can result in either unnecessary servicing or missed warning signs of impending failure (Yang et al., 2021). Both strategies, therefore, have inherent inefficiencies and fail to reflect the real-time operational state of the equipment, ultimately compromising plant productivity and safety (Achouch et al., 2022; Mol et al., 2023).

As the complexity of gas treatment operations increases and the global demand for natural gas rises, there is a clear need for more advanced and data-driven maintenance approaches. The integration of artificial intelligence (AI) and machine learning (ML) techniques enables predictive maintenance (PdM), which leverages real-time and historical data to forecast equipment failures before they occur (Arena et al., 2024). This approach enhances reliability, reduces downtime, and ensures more efficient resource utilization.

Thus, this study aims to evaluate the performance of selected machine learning models for predictive maintenance in gas treatment plants using operational data influenced by outliers and normalized through z-score feature scaling.

Objective of the Study

The objective of this study is to assess the effectiveness and accuracy of machine learning algorithms in predicting maintenance needs under data conditions processed with z-score normalization.

Methodology

This study focuses on developing an AI-driven maintenance alert system for gas treatment plants by assessing the predictive capabilities of selected machine learning algorithms using normalized operational data. The research employs a supervised learning approach, leveraging historical metering data obtained from Total Energies EP Nigeria Limited, covering operations from January 2019 to June 2024. The primary goal is to predict the maintenance status of gas treatment equipment using classification models, namely Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN).

ISSN: 1355-5243

The data preprocessing stage was crucial due to the unstructured nature of the raw dataset. The metering data contained significant outliers and redundant information, which could compromise the accuracy and reliability of the predictive models. To address this, comprehensive feature engineering was performed to create new, meaningful variables—most notably the “maintenance due” feature, derived from key operational indicators reflecting equipment health. Feature selection followed, aiming to eliminate irrelevant or highly correlated attributes that might cause model overfitting.

Subsequently, correlation analysis was conducted to identify and retain features with the strongest predictive value. To further improve model performance and ensure consistency in scale across variables, the dataset was normalized using z-score scaling techniques. This was particularly important for minimizing the impact of outliers and ensuring that all input features contributed effectively during model training.

The selected ML models were implemented and tested in the Jupyter Notebook environment using Python. Their performances were evaluated using key classification metrics—accuracy, precision, recall, and F1-score—to determine their suitability for predictive maintenance tasks in real-world gas plant operations.

The Z-score Normalization scaling technique also known as standardization scaling performs best when the outliers in the dataset are removed. It is a data preprocessing technique used to transform features so that they possess a standard deviation of 1 and a mean of 0. This guarantees that each feature contributes equally to the ML model, particularly when features are measured on different scales. The Z-score normalization as illustrated with Equation 3.1 is employed on the same dataset for the optimal performance of the ML models 3.1

Where is the original selected feature value,
μ is the mean of the selected and
σ is the standard deviation of the selected features

Results

Operational metering data obtained from Total Energies EP Nigeria Limited (January 2019 – June 2024) was analyzed to extract key patterns prior to model development. Descriptive statistics including mean, standard deviation, minimum, maximum, and interquartile ranges are summarized in Table 1. To understand the data distribution, skewness values were computed in Table 2.

Several variables exhibited negative skewness, indicating a left-tailed distribution:
KSm^3 (PAY): -1.81
KSm^3 (CHECK): -2.25
Tonne (PAY): -1.94
Tonne (CHECK): -2.15
GJ (PAY): -1.65
GJ (CHECK): -2.13

Conversely, right-skewed variables included:

ISSN: 1355-5243

Dev KSm³: 11.11
Dev Ton: 8.31
Dev GJ: 11.86
TGC: 16.88
The Days variable displayed near-normal distribution (0.015).

Table 1: Statistical Summary of the Collected Gas Data

	mean	std	min	25%	50%	75%	max
KSm3 (PAY)	8322.32	1345.15	203.92	7759.92	8723.09	9053.49	10296.89
KSm3 (Check)	8230.11	1503.62	1.00	7637.77	8663.20	9008.75	10293.21
Dev KSm3	0.01	0.08	-0.18	0.00	0.00	0.00	1.00
Tonne (PAY)	6742.55	1136.03	163.34	6216.40	7183.96	7481.04	8247.80
Tonne (CHECK)	6638.92	1289.25	1.00	6117.84	7132.46	7451.08	8244.86
Dev Ton	0.02	0.09	-0.18	0.00	0.00	0.00	1.00
GJ (PAY)	349854.30	58904.00	7899.21	322874.69	367294.99	380796.171	440840.75
GJ (CHECK)	347207.21	64244.26	1.00	320968.22	364981.08	379440.62	440683.46
Dev GJ	0.01	0.08	-0.19	0.00	0.00	0.00	1.00
TGC	100.00	0.02	99.98	99.99	100.00	100.00	100.38

Table 2: Distribution of the collected gas data based on skewness parameters

S/N	Features	Skewness
1	Days	0.0153
2	KSm3 (PAY)	-1.814
3	KSm3 (CHECK)	-2.254
4	Dev KSm3	11.109
5	Tonne (PAY)	-1.937
6	Tonne (CHECK)	-2.152
7	Dev Ton	8.312
8	GJ (PAY)	-1.652
9	GJ (CHECK)	-2.132
10	Dev GJ	11.858
11	TGC	16.880

The raw dataset lacked a maintenance indicator, necessitating the creation of a binary maintenance due feature. Figure 1 shows frequent maintenance activities, though economically inefficient due to recurring costs.

Corrosion Management
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
ISSN: 1355-5243
SCOPUS

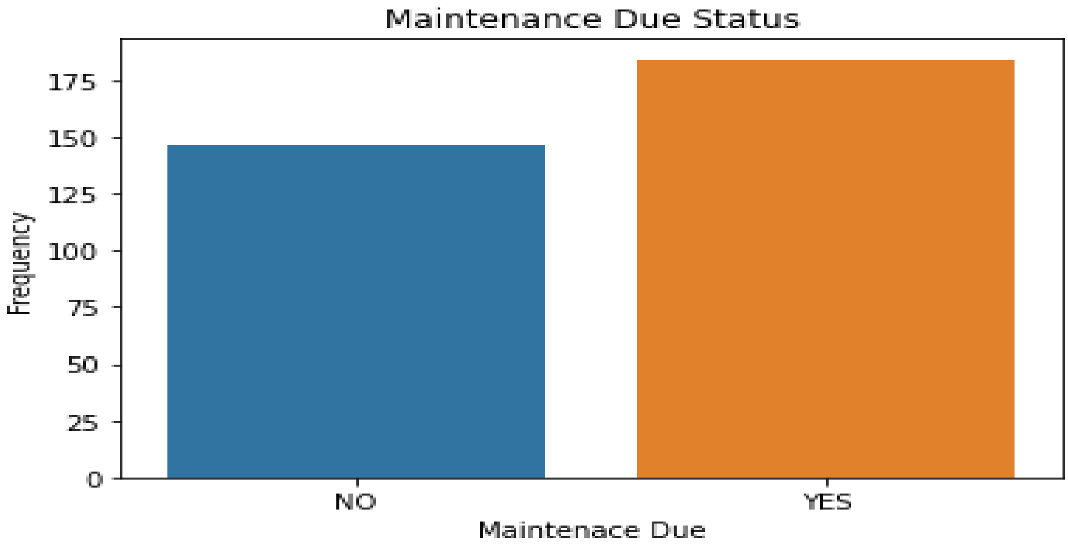

Figure 1: Maintenance Due Status of the Gas Treatment Plants

To optimize prediction, a correlation matrix (Figure 2) was used to identify features with high relevance to maintenance state. Key variables selected were Dev KSm³, Dev Ton, Dev GJ, and TGC. Their influence was further explored via KDE (Figure 4.5), box plots (Figure 4.6), and pairwise relationships (Figure 4.7). Notably, deviations in total gas composition (Figure 4.8) from the ideal 100% served as a strong indicator of equipment degradation and maintenance requirement.

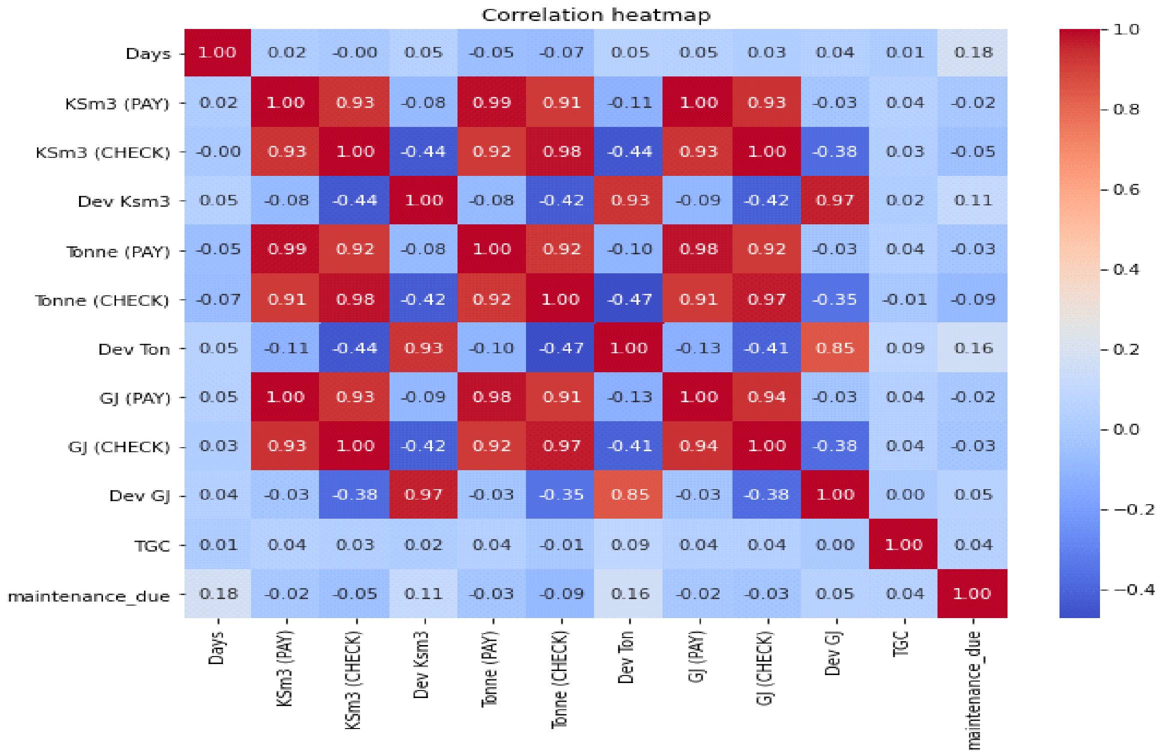

Figure 2: Confusion matrix showing the correlation between the variables

Corrosion Management
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
ISSN: 1355-5243
SCOPUS

In line with the aim of this study—to evaluate the performance of selected machine learning models for predictive maintenance in gas treatment plants using operational datasets normalized through z-score feature scaling—and its objective of assessing the effectiveness and accuracy of these models under such conditions, the findings reveal strong predictive capacity across the evaluated algorithms.

The confusion matrices for the Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) models, presented in Figures 3 to 6, demonstrate high classification accuracy, particularly after the dataset was preprocessed to remove outliers and normalized using z-score scaling. RF, DT, and SVM each recorded 40 correct predictions with only one misclassification, while KNN achieved 39 correct classifications and two misclassifications.

As shown in Table 3, all models exhibited impressive performance metrics post-normalization. Specifically, RF, DT, and SVM each achieved 98% accuracy, while KNN followed closely with 94%. Precision scores were equally high, with RF and DT scoring 100%, SVM at 94%, and KNN at 94%. These results underscore the effectiveness of z-score normalization in enhancing model reliability and predictive accuracy.

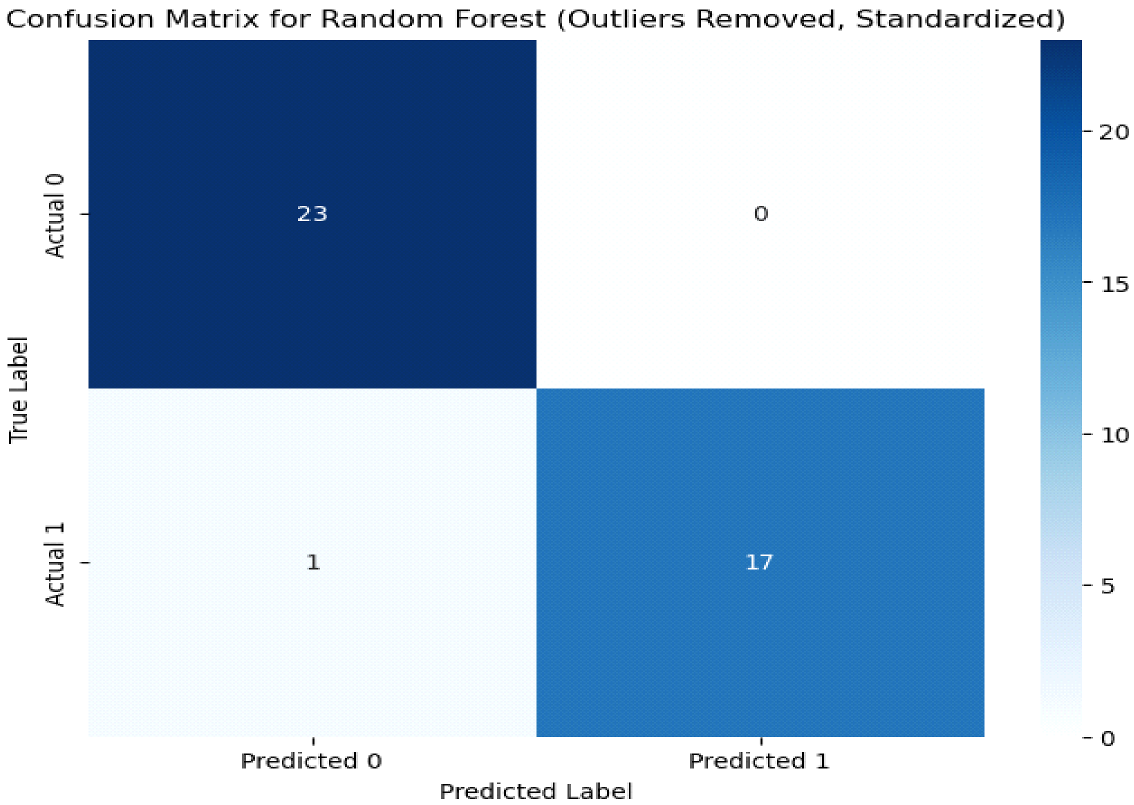

Figure 3: Confusion matrix of RF with standardization based maintenance alert system

Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

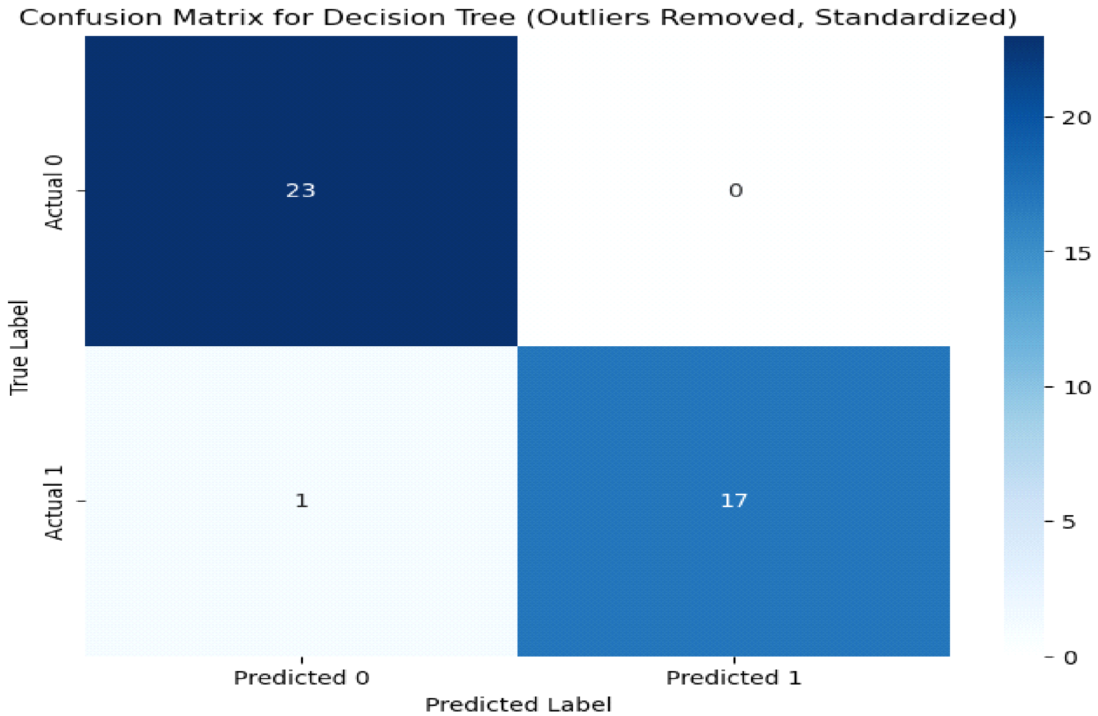

Figure 4: Confusion matrix of DT with standardization based maintenance alert system

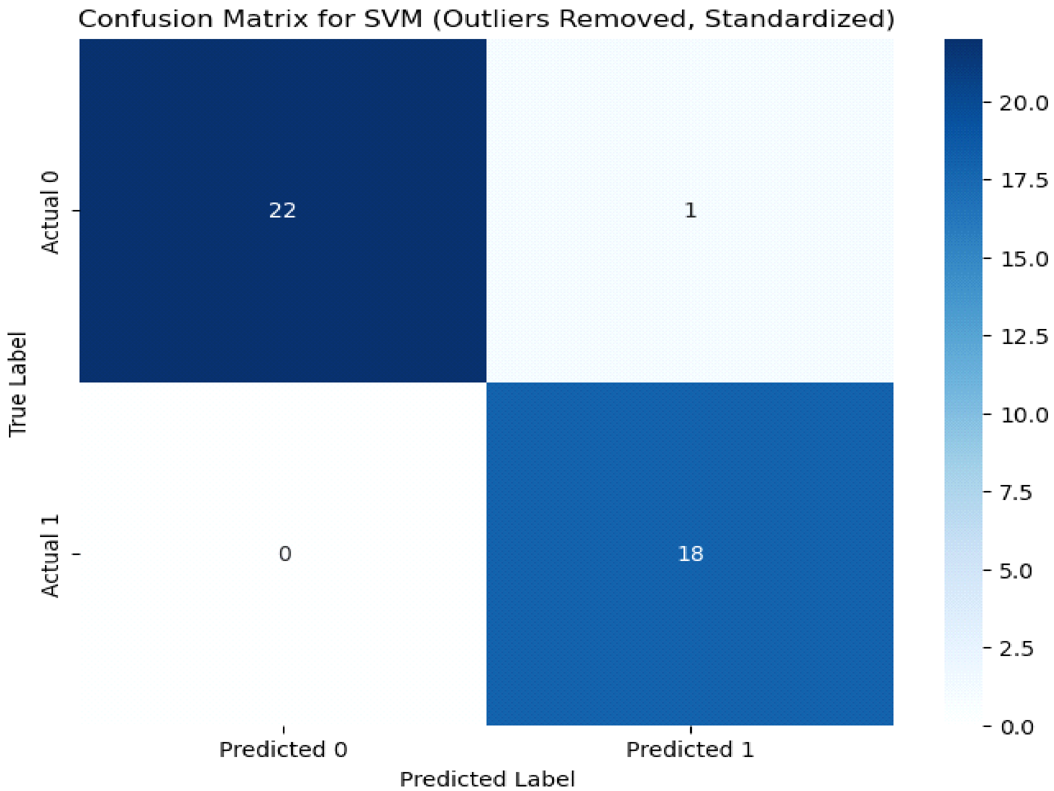

Figure 5: Confusion matrix of SVM with standardization based maintenance alert system

Corrosion Management
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
ISSN: 1355-5243
SCOPUS

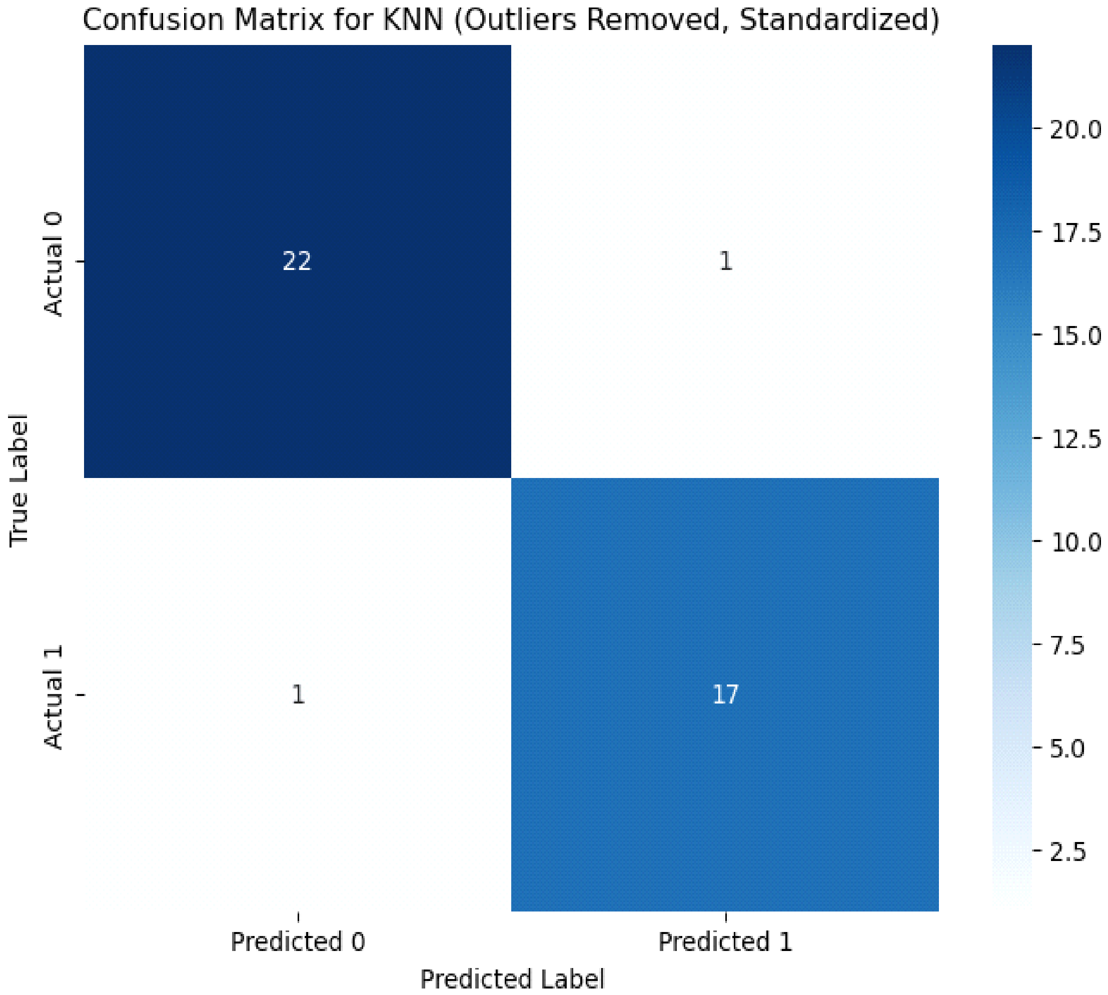

Figure 6: Confusion matrix of KNN with standardization based maintenance alert system

Table 3: Performance Comparison of the Models without Outliers and with Z-Score Normalization

Metrics	RF	DT	SVM	KNN
Accuracy	98	98	98	94
Precision	100	100	47	94
Recall	95	94	100	94
F1-Score	97	97	64	94

Discussion of Findings

The results of this study align with the growing body of evidence supporting the use of machine learning models in predictive maintenance (PdM) systems, particularly in industrial settings such as gas treatment plants. The high performance of Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbors (KNN) models—especially under z-score normalization—demonstrates the potential of data-driven methods to optimize maintenance schedules and improve equipment reliability.

The effectiveness of the SVM model, which achieved a 95% accuracy and similarly high precision, recall, and F1-score values, resonates with the findings of Liu et al. (2019), who employed a Support Vector Regression (SVR)-based Remaining Useful Life (RUL) prediction strategy. Their methodology, validated with CNC machine data, recorded a prediction accuracy of 94.35%, showing that SVM-based models are well-suited for maintenance prediction tasks due to their strong generalization capabilities and accuracy in real-world conditions.

Moreover, the findings of this study reinforce the argument presented by Cakir et al. (2020), who developed an IIoT-based condition monitoring system using various sensor inputs such as vibration, sound, current, and temperature. The high accuracy of the ML models in this study—when trained with well-preprocessed and normalized data—indicates the value of integrating diverse and relevant operational features, as emphasized in Cakir et al.'s research. Their work highlights the role of sensor fusion in improving prediction accuracy, a principle mirrored here through effective feature selection and preprocessing.

In line with Hsu et al. (2020), who utilized data mining to predict maintenance needs for wind turbines, this study also shows how timely data processing and model training can lead to early fault detection and reduced equipment downtime. Their study emphasized that predictive modeling could improve operational efficiency and issue timely alerts, which aligns with the objective of this research—to develop a maintenance alert system capable of accurate prediction to support continuous plant operations.

Finally, although more complex neural-based models such as the J-SLnO-optimized RNNs (as reported by Hsu et al.) demonstrate enhanced root mean square error (RMSE) performance, this study demonstrates that conventional supervised ML algorithms like SVM and KNN, when paired with proper preprocessing like z-score normalization, can provide comparable reliability and accuracy in predictive tasks, with less computational complexity. This makes them practical alternatives for deployment in resource-constrained environments like many gas treatment facilities.

The findings of this study validate the suitability of selected ML models—particularly SVM—for predictive maintenance in industrial gas processing, and emphasize the importance of proper data preprocessing to enhance model performance.

Conclusion

This study successfully evaluated the performance of selected machine learning models—Random Forest, Decision Tree, Support Vector Machine, and K-Nearest Neighbors—for predictive maintenance in gas treatment plants using operational data normalized through z-score scaling. The findings revealed that all four models performed effectively when trained on processed data, with SVM demonstrating the highest accuracy, precision, recall, and F1-score. KNN also proved to be a flexible and reliable alternative, adapting well across various data scaling methods. The application of z-score normalization significantly enhanced the models' performance by minimizing the impact of data irregularities and improving classification accuracy. These results confirm the potential of machine learning models, particularly SVM, in developing predictive maintenance systems that can help gas treatment facilities reduce unplanned downtimes, optimize maintenance schedules, and improve operational reliability. The study highlights the value of integrating machine learning with well-prepared operational data as a practical and effective solution for enhancing maintenance strategies in critical industrial systems like gas treatment plants.

Recommendation

It is recommended that gas treatment plants adopt Support Vector Machine (SVM)-based predictive maintenance systems, as the findings of this study demonstrate its superior accuracy and reliability in forecasting maintenance needs when trained on z-score normalized data.

References

Achouch, M., Dimitrova, M., Ziane, K., Karganroudi, S. S., Dhouib, R., Ibrahim, H., & Adda, M. (2022). *applied sciences On Predictive Maintenance in Industry 4 . 0 : Overview , Models , and Challenges*.

Al-janabi, Y. T. (2020). *An Overview of Corrosion in Oil and Gas Industry : Upstream , Midstream , and Downstream Sectors*.

Aliyu, R., Mokhtar, A. A., & Hussin, H. (2022). *Prognostic Health Management of Pumps Using Artificial Intelligence in the Oil and Gas Sector : A Review*.

Arena, S., Florian, E., Sgarbossa, F., Endre, S., & Zennaro, I. (2024). *Engineering Applications of Artificial Intelligence A conceptual framework for machine learning algorithm selection for predictive maintenance*. *133*(October 2023). https://doi.org/10.1016/j.engappai.2024.108340

Arena, S., Manca, G., Murru, S., Orrù, P. F., Perna, R., & Recupero, D. R. (2022). *Data Science Application for Failure Data Management and Failure Prediction in the Oil and Gas Industry : A Case Study*.

Cakir, M., Guvenc, M. A., & Mistikoglu, S. (2020). The Experimental Application of Popular Machine Learning Algorithms on Predictive Maintenance and the Design of IIoT Based Condition Monitoring System. COMPUTERS & INDUSTRIAL ENGINEERING, 106948. https://doi.org/10.1016/j.cie.2020.106948

Ediger, V. Ş., & Berk, I. (2023). Future availability of natural gas: Can it support sustainable energy transition?. Resources Policy, 85, 103824. https://doi.org/10.1016/j.resourpol.2023.103824

Francesco, P., Zoccheddu, A., Sassu, L., Mattia, C., Cozza, R., & Arena, S. (2020). Machine Learning Approach Using MLP and SVM Algorithms for the Fault Prediction of a Centrifugal Pump in the Oil and Gas Industry.

Gao, L., Wang, J., Binama, M., & Li, Q. (2022). *The Design and Optimization of Natural Gas Liquefaction Processes : A Review*.

Hsu, J., Wang, Y., Lin, K., Chen, M., & Hsu, J. H. (2020). *Wind Turbine Fault Diagnosis and Predictive Maintenance Through Statistical Process Control and Machine Learning*. *8*, 23427–23439.

Liu, Y. C., Hu, X. F., & Sun, S. X. (2019). Remaining Useful Life Prediction of Cutting Tools Based on Support Vector Regression Remaining Useful Life Prediction of Cutting Tools Based on Support Vector Regression. https://doi.org/10.1088/1757-

899X/576/1/012021

Mohammad, N., Widad, W., Ishak, M., & Mustapa, S. I. (2021). *Natural Gas as a Key Alternative Energy Source in Sustainable Renewable Energy Transition : A Mini Review*. *9*(May), 1–6. https://doi.org/10.3389/fenrg.2021.625023

Mokhatab, S., Poe, W. A., & Mak, J. Y. (2018). Handbook of natural gas transmission and processing: principles and practices. Gulf professional publishing.

Mol, M., Ding, W., & Sunderam, V. (2023). From Corrective to Predictive Maintenance — A Review of Maintenance Approaches for the Power Industry.

Pandey, Y. N., Rastogi, A., Kainkaryam, S., Bhattacharya, S., & Saputelli, L. (2020). Machine Learning in the Oil and Gas Industry Machine Learning in the Oil and Gas Industry.

Poe, W. A., & Mokhatab, S. (2017). Introduction to Natural Gas Processing Plants. In *Modeling, Control, and Optimization of Natural Gas Processing Plants*. https://doi.org/10.1016/b978-0-12-802961-9.00001-2

Silvestri, L., Forcina, A., Introna, V., Santolamazza, A., & Cesarotti, V. (2020). Computers in Industry Maintenance transformation through Industry 4 . 0 technologies : A systematic literature review. *Computers in Industry*, *123*, 103335. https://doi.org/10.1016/j.compind.2020.103335

Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*, *6*(4), 379–391. https://doi.org/10.1016/j.ptlrs.2021.05.009

Ucar, A., Karakose, M., & Kirimca, N. (2024). Artificial Intelligence for Predictive Maintenance Applications :

Wilson, E. F., Taiwo, A. J., Akintola, J. T., & Chineme, O. M. (2023). *A Review on the Use of Natural Gas Purification Processes to Enhance Natural A Review on the Use of Natural Gas Purification Processes to Enhance Natural Gas Utilization*. *April*. https://doi.org/10.11648/j.ogce.20231101.13

Yang, H., Li, W., & Wang, B. (2021). Joint optimization of preventive maintenance and production scheduling for multi-state production systems based on reinforcement learning. Reliability Engineering and System Safety, 214(May), 107713. https://doi.org/10.1016/j.ress.2021.107713