MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 01-14

Robust Machine Learning Framework for Predictive Maintenance in Gas Treatment Plants under Outlier-Influenced Operational Conditions

Femi Adeoye Alabi[1]
Department of Electrical/Electronic Engineering, Bells University of Technology, Ota, Ogun State
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

Abraham Olatide Amole[3]
Department of Electrical/Electronic Engineering, Bells University of Technology, Ota, Ogun State
aoamole@bellsuniversity.edu.ng

Abstract

The increasing complexity of gas treatment plants (GTPs) and the rising global demand for natural gas underscore the need for efficient and reliable maintenance strategies. Traditional maintenance approaches—reactive and preventive—have shown limitations in predicting failures accurately, often resulting in costly downtime and inefficient resource utilization. This study aims 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)—in predicting maintenance requirements in GTPs under data conditions involving outliers and the application of robust scaling techniques. Operational metering data from January 2019 to June 2024 were obtained from Total Energies EP Nigeria Limited. After preprocessing, feature engineering and selection were performed to prepare the dataset for training the models in Python's Jupyter Notebook environment. The models were assessed based on accuracy, precision, recall, and F1-score. Results revealed that RF and DT models achieved an ideal 100% accuracy even in the presence of outliers and after applying robust feature scaling, although this level of perfection is unlikely in real-world conditions. KNN showed moderate performance with an accuracy of 89%, while SVM performed poorly with an accuracy of 47%, indicating its high sensitivity to outliers despite scaling. The findings confirm that tree-based algorithms such as RF and DT are more resilient and suitable for predictive maintenance tasks in noisy industrial datasets. Therefore, the study concludes that integrating robust scaling with tree-based ML models provides a more dependable solution for predictive maintenance in GTPs. It is recommended that GTPs adopt RF or DT models alongside robust data preprocessing techniques to improve fault prediction, reduce unplanned downtimes, and enhance operational reliability.

Keywords: Predictive Maintenance, Gas Treatment Plants, Machine Learning, Outliers, Robust Scaling.

Introduction

Energy serves as the foundation of modern civilization, playing a vital role in driving economic progress, technological innovation, and overall societal development. Over the last century, global energy consumption has risen significantly, reaching 552 quadrillion British thermal units (BTUs) in 2016 due to factors such as population expansion, urbanization, and growing technological dependence (Pandey et al., 2020). In that same year, the oil and gas (O&G) sector accounted for 55% of the global energy supply, a figure expected to increase to 57% by 2040 (Ediger et al., 2023). Among the various fossil fuels, natural gas has gained preference due to its cleaner combustion characteristics, versatility, and availability. Its lower carbon emissions, especially when compared to coal, make it a more environmentally sustainable option for electricity generation and industrial use (Gao et al., 2022; Mohammad et al., 2021).

Processing natural gas involves a complex infrastructure—from extraction and transportation to purification at gas treatment plants (GTPs). GTPs are crucial facilities that ensure the natural gas meets safety and quality standards by removing harmful contaminants such as hydrogen sulfide (H_2S), carbon dioxide (CO_2), and water vapor (Mokhatab et al., 2018). These impurities can compromise gas quality, corrode pipelines, and cause equipment failure. Therefore, efficient operation of GTPs is essential for reliable gas delivery and energy security (Wilson et al., 2023).

Equipment in GTPs—including compressors, separators, and heat exchangers—is exposed to extreme conditions like high pressure, elevated temperatures, and corrosive gases, which accelerate wear and tear (Poe & Mokhatab, 2017). Al-janabi (2020) noted that the presence of hydrogen sulfide, in particular, significantly shortens equipment lifespan and raises maintenance costs. Although regular maintenance is essential for ensuring equipment efficiency and safety (Arena et al., 2022), the unpredictable nature of equipment degradation often leads to sudden failures, causing operational delays and financial losses (Theissler et al., 2021).

Historically, GTPs have relied on reactive and preventive maintenance strategies. Reactive maintenance involves addressing equipment faults only after failure has occurred, leading to prolonged downtimes and increased repair costs (Ucar et al., 2024). Preventive maintenance, on the other hand, is based on scheduled inspections and historical trends, aiming to reduce the frequency of failures (Cheng et al., 2020; Achouch et al., 2022). However, preventive maintenance is not always efficient, as it can lead to unnecessary servicing, resource wastage, and misalignment with the actual equipment condition (Yang et al., 2021). Both strategies are costly, often contributing to over one-third of operational expenditures (Francesco et al., 2020).

In response to these limitations, predictive maintenance (PdM) has emerged as a more effective alternative. It utilizes real-time data and historical records to forecast potential failures before they occur, allowing timely interventions and minimizing disruptions (Ucar et al., 2024). PdM strategies are increasingly supported by Industry 4.0 technologies, including Internet of Things (IoT) sensors and intelligent data processing systems (Silvestri et al., 2020). Through the application of artificial intelligence (AI) and machine learning (ML), large volumes of sensor data can be analyzed to detect subtle signs of equipment deterioration (Arena et al., 2024).

Recent studies have demonstrated the effectiveness of ML techniques—such as Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbors (KNN)—in predicting equipment health and preventing failures in the O&G industry (Sircar et al., 2021). Aliyu et al. (2022) emphasized the value of these models in detecting faults in pump systems, enabling proactive maintenance decisions. However, one major challenge in deploying ML models in industrial environments is the presence of noisy data and outliers, which can significantly distort model predictions. Applying robust feature scaling can help minimize the influence of such irregularities and enhance model reliability.

This study aims to assess 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 datasets containing outliers and processed through robust feature scaling, thereby addressing real-world data challenges and supporting more accurate maintenance planning.

Statement of the Study

Gas Treatment Plants (GTPs) are vital facilities in the oil and gas industry, responsible for ensuring the delivery of high-quality, purified natural gas. However, the harsh operational environments in which these plants function—marked by high pressure, extreme temperatures, and exposure to corrosive substances—result in periodic wear and tear of critical components such as compressors, separators, heat exchangers, and scrubbers (Poe & Mokhatab, 2017). Ensuring the continuous and reliable performance of this equipment is essential to meet rising energy demands. Unfortunately, traditional maintenance strategies such as reactive maintenance (RM) and preventive maintenance (PM) have proven inadequate in addressing these challenges effectively.

Reactive maintenance is only triggered after equipment failure has occurred, leading to prolonged downtimes and high repair costs (Abidi et al., 2022). Studies by Achouch et al. (2022) and Mol et al. (2023) confirm that RM can increase downtime by up to 50% due to the unexpected nature of failures. On the other hand, while preventive maintenance is more proactive, it is still limited by its dependence on historical data and rigid scheduling, often leading to unnecessary maintenance interventions that waste time and resources (Yang et al., 2021).

Given these constraints, there is a growing need for intelligent, data-driven strategies that can accurately predict equipment failures before they occur. Predictive maintenance (PdM), powered by artificial intelligence (AI) and machine learning (ML), presents a promising solution by analyzing real-time operational data to detect early signs of equipment degradation (Arena et al., 2024). This approach enhances maintenance planning, reduces costs, and improves the reliability and safety of GTP operations.

In light of this, the present study is focused on developing an AI-driven maintenance alert system capable of predicting equipment maintenance needs using supervised machine learning models. By integrating robust data preprocessing techniques and evaluating model performance under realistic industrial conditions—including the presence of outliers and the application of robust scaling—this study aims to provide a reliable and efficient maintenance solution tailored to the evolving complexities of gas treatment plants.

Objective of the Study

The objective of this study is to assess the effectiveness and resilience of machine learning algorithms in predicting maintenance needs under data conditions involving outliers and the application of robust scaling techniques.

Methodology

This study is centered on the development of an intelligent maintenance alert system for gas treatment plants using machine learning techniques. The approach involves utilizing historical operational metering data obtained from Total Energies EP Nigeria Limited, covering the period from January 2019 to June 2024. The dataset comprises multiple spreadsheets containing unstructured plant metering data, which required careful preprocessing before model training. Four supervised classification algorithms—Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)—were selected to predict the maintenance status of the plant equipment. These algorithms were implemented and simulated within the Jupyter Notebook Python environment to assess their capability in forecasting when maintenance is due. The preprocessing phase involved cleaning, organizing, and transforming the raw metering data into a structured format suitable for analysis. Feature engineering was carried out to create relevant variables, including a target feature—'maintenance due'—derived from specific operational indicators reflecting the condition of the equipment. Redundant, irrelevant, or highly correlated features were eliminated through feature selection to improve model efficiency and avoid overfitting. Correlation analysis was employed to retain the most informative features for model training. Due to the presence of significant outliers in the dataset, appropriate feature scaling techniques were applied to normalize the data. Both z-score normalization and robust scaling were used to minimize the impact of extreme values and enhance model performance. The trained models were evaluated using performance metrics such as accuracy, precision, recall, and F1-score to determine their effectiveness in predicting maintenance needs under realistic plant conditions. A flowchart summarizing the methodological process is presented in Figure 1.

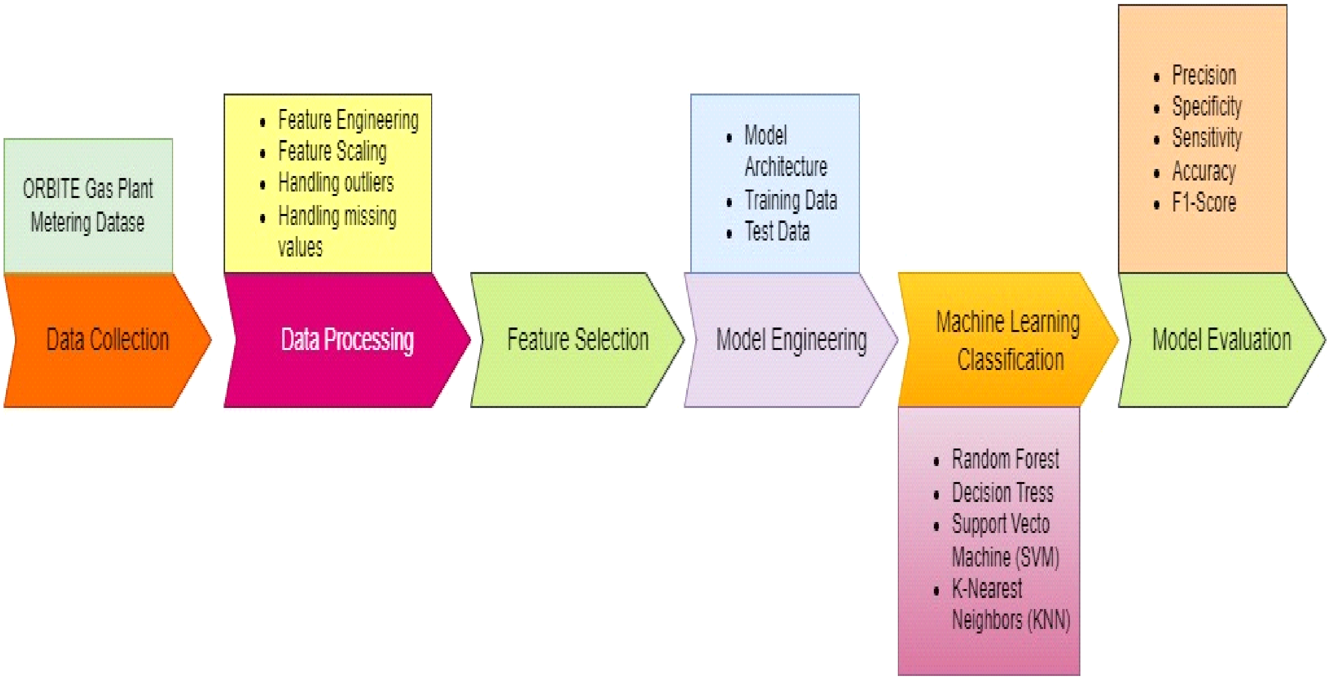

Figure 1: Overall Procedure Diagram for the Research

Robust scaling is applied to the data without removing the outliers to observe the performance of the developed models. Robust scaling is a data normalization technique used to transform features in a dataset so they are less influenced by outliers or extreme values. This scaling technique uses the median and interquartile range to scale the selected features making it robust to the outliers as illustrated in Equation 3.2

3.2

Where is the original selected feature value,
 is the median of the selected value,
 is the interquartile range where Q1 is the 25th percentile or lower quartile and Q3 is the 75th percentile or upper quartile.
The interquartile range represents the spread of the middle 50% of the data, making it more resistant to the effects of outliers.

Results

Operational metering data gathered from gas treatment facilities operated by Total Energies EP Nigeria Limited between January 2019 and June 2024 was first subjected to exploratory data analysis (EDA). This initial phase was crucial to uncover patterns and better understand the structure and behavior of the dataset prior to model training. Summary statistics including the mean, standard deviation, minimum, maximum, and percentiles (25th, 50th, and 75th) were calculated for each variable, as presented in Table 1.

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

To ensure robust and unbiased model outcomes, the distribution of each feature was examined. Table 2 highlights the skewness values, providing insight into the symmetry—or lack thereof—of the data distribution:

KSm^3 (PAY): -1.814
KSm^3 (CHECK): -2.254
Tonne (PAY): -1.937
Tonne (CHECK): -2.152
GJ (PAY): -1.652
GJ (CHECK): -2.132
On the other hand, some variables showed considerable rightward skewness:
Dev KSm^3: 11.109
Dev Ton: 8.312
Dev GJ: 11.858
TGC: 16.880

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

Only the Days variable maintained near-perfect symmetry, with a skewness of 0.015, indicating normal distribution.

Since the original dataset did not contain a predefined indicator for maintenance requirements, a derived variable—maintenance_due—was introduced to represent when the equipment is likely in need of servicing.

To improve the efficiency of maintenance forecasting, correlation analysis was conducted between the available features and the maintenance_due variable (see Figure 2). From this assessment, four key variables were identified as being significantly related to maintenance needs:
Dev KSm³
Dev Ton
Dev GJ
Total Gas Composition (TGC)

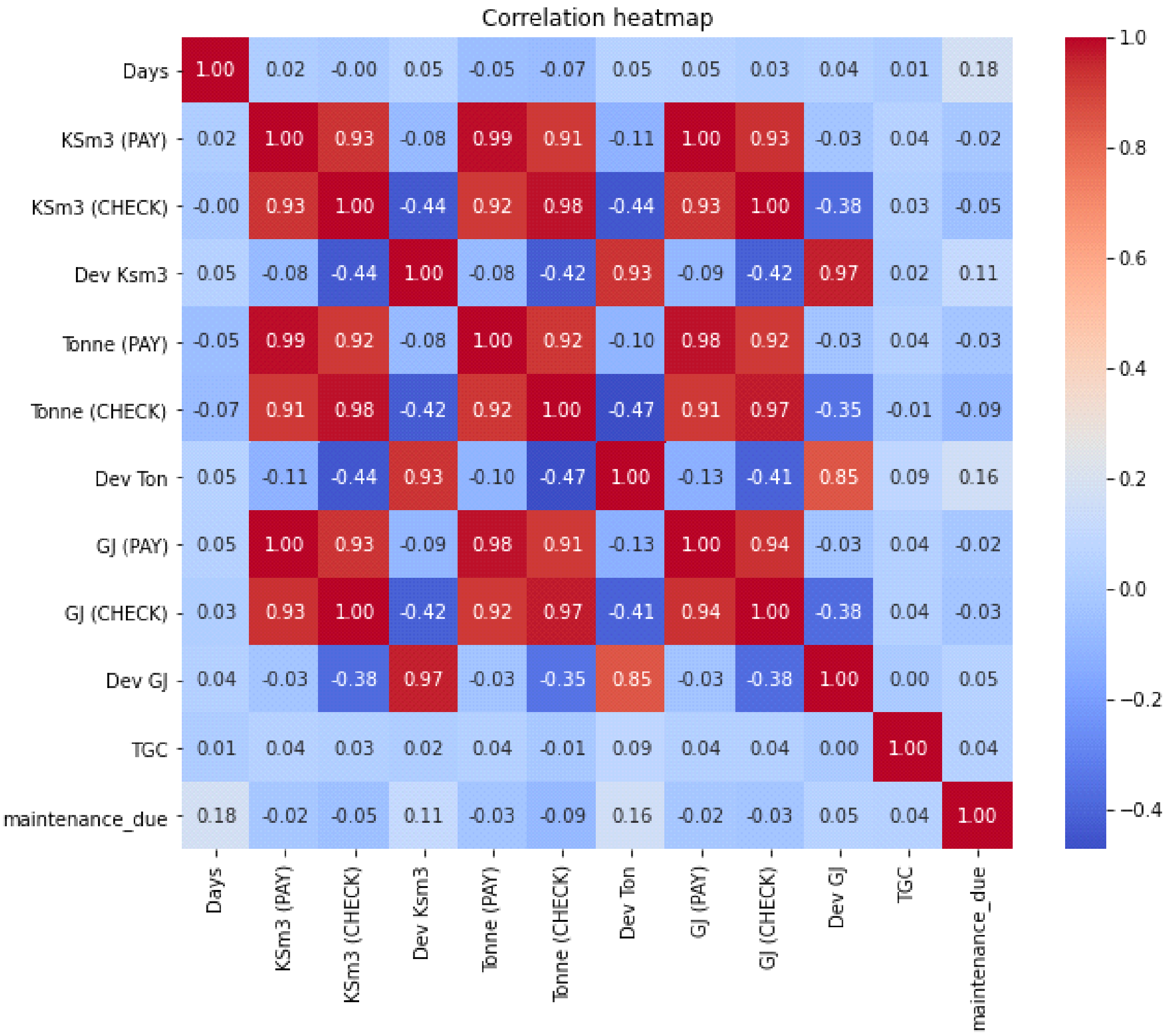

Figure 2: Confusion matrix showing the correlation between the variables

The Total Gas Composition (TGC), in particular, stood out as a vital indicator. Ideally, the combined composition of gases should sum up to 100%. Any deviation from this target signals potential imbalance or system inefficiency, prompting the need for maintenance.

This study set out to evaluate the performance of selected machine learning models—Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)—for predictive maintenance in gas treatment plants using datasets containing outliers and processed through robust scaling techniques.

The confusion matrices presented in Figures 3 to 6 highlight the classification outcomes for each algorithm under these conditions. RF and DT models both achieved perfect classification results, correctly identifying all 66 instances as either "maintenance due" or "maintenance not due," with no misclassifications. Conversely, SVM misclassified 35 out of 66 instances, correctly classifying only 31, indicating poor performance under the influence of outliers despite the use of robust scaling. KNN achieved 59 correct classifications and 7 misclassifications, showing improved performance compared to SVM but still less accurate than RF and DT.

Table 3 summarizes the performance metrics—accuracy, precision, recall, and F1-score—used to assess each model. RF and DT attained 100% accuracy, which, although indicative of strong performance, may be too ideal for real-world deployment, raising concerns about possible overfitting. KNN followed with an accuracy of 89%, making it a more realistic yet moderately effective option. In contrast, SVM exhibited the weakest performance with an accuracy of just 47%, reinforcing its sensitivity to outliers and lower robustness, even when scaling was applied.

From the results, it can be inferred that while RF and DT models are highly effective under the given conditions, their overly perfect results suggest the need for cautious validation. KNN shows reasonable reliability, while SVM underperforms and is unsuitable for predictive maintenance tasks in gas treatment environments with outlier-influenced datasets. These findings align with the study's objective to assess the robustness of ML models under imperfect real-world data conditions.

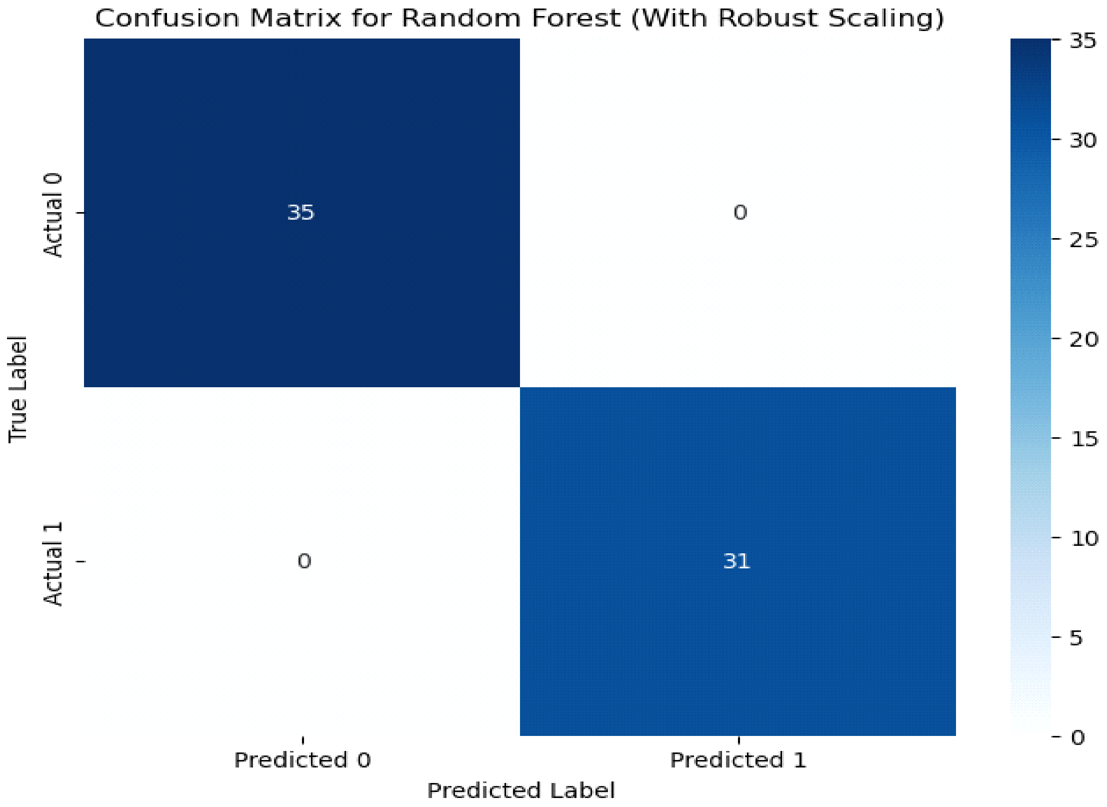

Figure 3: Confusion matrix of RF with robust scaling based maintenance alert system

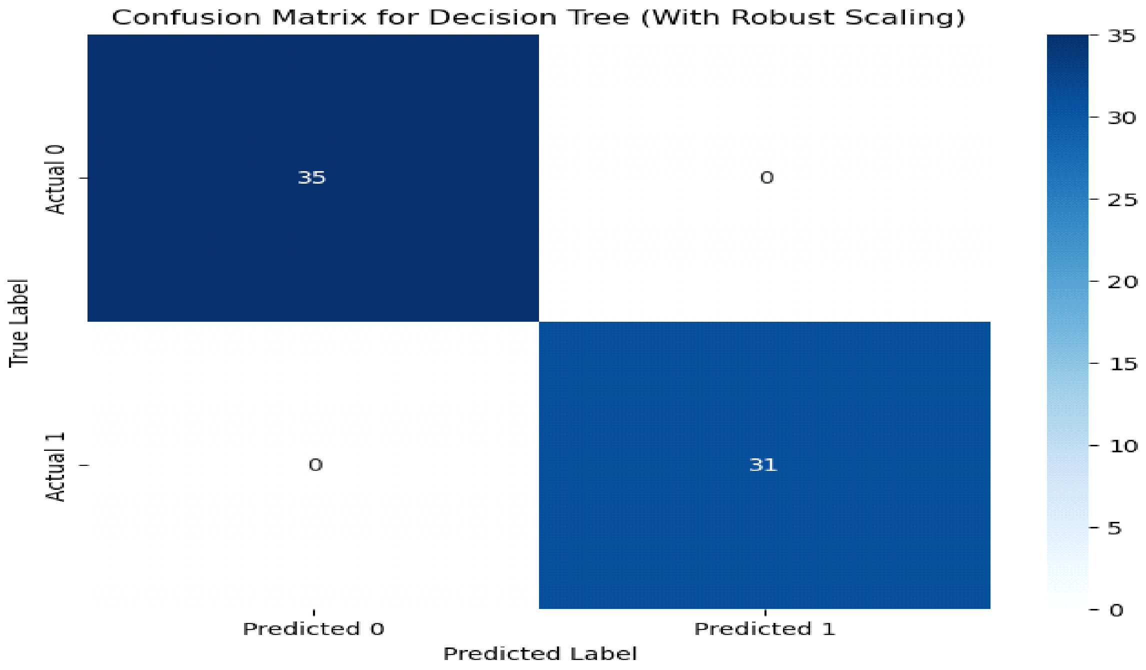

Figure 4: Confusion matrix of DT with robust scaling based maintenance alert system

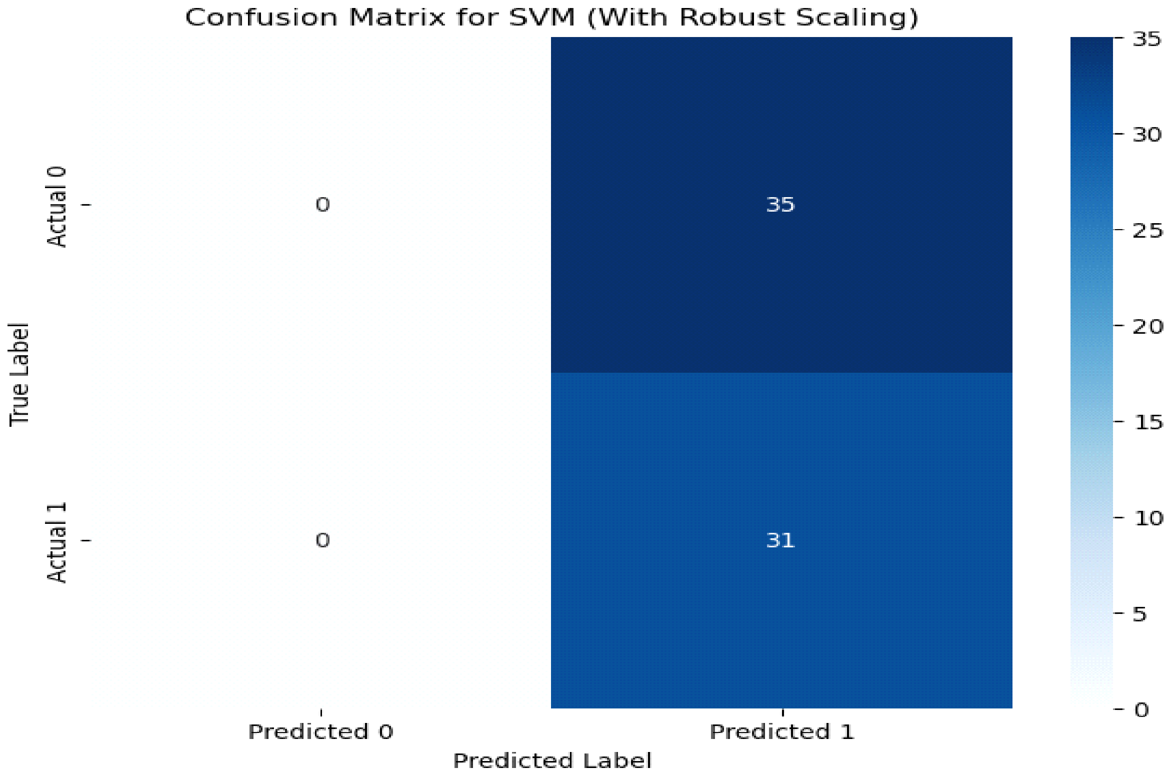

Figure 5: Confusion matrix of SVM with robust scaling based maintenance alert system

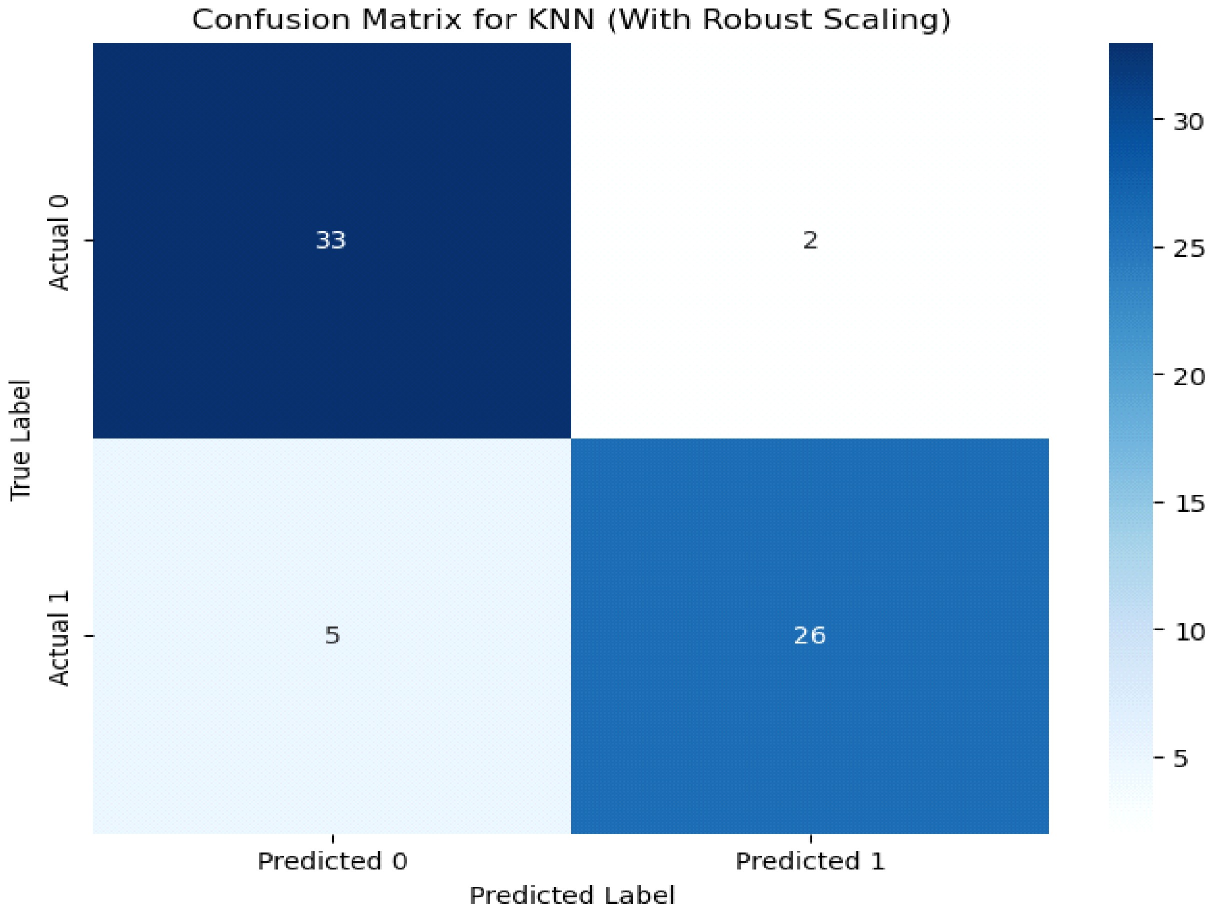

Figure 6: Confusion matrix of KNN with robust scaling based maintenance alert system

Table 3: Performance Comparison of the Models with Outliers and With Robust Scaling

Metrics	RF	DT	SVM	KNN
Accuracy	100	100	47	89
Precision	100	100	47	93
Recall	100	100	100	84
F1-Score	100	100	64	88

Discussion of Findings

The findings of this study reveal significant insights into the predictive maintenance of gas treatment plants using machine learning (ML) techniques under conditions involving outliers and robust feature scaling. Among the four evaluated algorithms—Random Forest (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM)—RF and DT showed perfect classification accuracy (100%) in predicting whether the gas plant was due or not due for maintenance. KNN followed with an 89% accuracy, while SVM significantly underperformed with an accuracy of only 47%. These results underscore the resilience of tree-based models (RF and DT) in handling noisy data, including outliers, even after the application of robust scaling, while highlighting the limitations of SVM in such scenarios.

These outcomes align with existing literature that explores the use of ML models for predictive maintenance in complex industrial environments. For instance, Makridis et al. (2021) presented an ML-based anomaly detection framework for vessel engines using models like OneClass SVM and LSTM. Their two-stage methodology—predicting expected values and comparing them to actual sensor readings—demonstrated effectiveness in identifying early signs of component failure. This aligns with the goals of this study, where early and accurate detection of maintenance needs is critical.

Similarly, T et al. (2021) demonstrated how ARIMA-based monitoring could reduce catastrophic failures by up to 84%, emphasizing the value of predictive insights in minimizing unplanned downtime. This supports the rationale for using machine learning to anticipate failures before they disrupt operations—an objective central to this study.

Furthermore, Zhou et al. (2019) effectively used LSTM models to estimate the remaining useful life (RUL) of tools, achieving low mean squared error values. This shows that deep learning models, when properly tuned and applied to structured datasets, can yield highly reliable forecasts. However, in this study, traditional ML models like RF and DT outperformed SVM, which supports the notion that model selection should be tailored to the data characteristics and domain-specific requirements.

The study affirms that while SVM may falter under real-world data irregularities, ensemble and tree-based models like RF and DT provide more consistent and robust performance. This strengthens the case for their integration into predictive maintenance systems for gas treatment plants, especially under operational conditions that often involve unstructured and noisy data.

Conclusion

This study has demonstrated the effectiveness of machine learning models in predicting maintenance needs for gas treatment plants, particularly under challenging data conditions involving outliers and the use of robust scaling techniques. Among the models evaluated—Random Forest, Decision Tree, K-Nearest Neighbors, and Support Vector Machine—the Random Forest and Decision Tree models delivered the highest accuracy, successfully classifying all instances without error. This highlights their robustness and suitability for real-world maintenance applications where data irregularities are common. Conversely, the Support Vector Machine showed significantly lower performance, indicating its sensitivity to noisy data even after preprocessing. The findings underscore the importance of selecting appropriate algorithms and preprocessing strategies in developing reliable predictive maintenance systems. By leveraging real-time operational data and applying suitable scaling methods, gas treatment facilities can improve equipment reliability, reduce unplanned downtime, and optimize maintenance schedules. Ultimately, the study affirms that tree-based machine learning models, when combined with effective data preprocessing, offer a practical and scalable solution for enhancing maintenance strategies in complex industrial environments.

Recommendation

Gas treatment plants should adopt tree-based machine learning models such as Random Forest and Decision Tree, combined with robust feature scaling techniques, to develop reliable predictive maintenance systems that minimize unplanned downtime and improve operational efficiency.

References

Abidi, M. H., Mohammed, M. K., & Alkhalefah, H. (2022). *Predictive Maintenance Planning for Industry 4 . 0 Using Machine Learning for Sustainable Manufacturing*.

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

Cheng, J. C. P., Chen, W., Chen, K., & Wang, Q. (2020). Automation in Construction Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. Automation in Construction, 112(January), 103087. https://doi.org/10.1016/j.autcon.2020.103087

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*.

Makridis, G., Kyriazis, D., & Plitsos, S. (2021). Predictive maintenance leveraging machine learning for time-series forecasting in the maritime industry. IEEE.

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

T, R. M., J, P. R., R, A. U., Devaraj, D., & Umachandran, K. (2021). Intelligent machine learning based total productive maintenance approach for achieving zero downtime in industrial machinery. Computers & Industrial Engineering, 157(March), 107267. https://doi.org/10.1016/j.cie.2021.107267

Theissler, A., Pérez-velázquez, J., Kettelgerdes, M., & Elger, G. (2021). Predictive maintenance enabled by machine learning : Use cases and challenges in the automotive industry. 215.

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

Zhou, J., Zhao, X., & Gao, J. (2019). *Tool remaining useful life prediction method based on LSTM under variable working conditions*.