Research Article GU J Sci, Part A, 12(1): 96-118 (2025) 10.54287/gujsa.1646993

Gazi University
Journal of Science
PART A: ENGINEERING AND INNOVATION
http://dergipark.org.tr/gujsa

Effective Maintenance of Industrial 5-Stage Compressor: A Machine Learning Approach

Okechukwu Chiedu EZEANYIM[1] Nnamdi Vitalis EWUZIE[1] Patrick Sunday AGUH[1] Chibuzo Victoria NWABUEZE[2] Charles Onyeka NWAMEKWE[1*]

[1] *Nnamdi Azikiwe University, Awka, Nigeria*
[2] *University of the People, Pasadena, USA*

Keywords

Cost-Effective Maintenance

Industrial Compressor

Machine Learning

Predictive Maintenance

Data-Driven Maintenance

Maintenance Optimization

Abstract

Effective maintenance is crucial in the manufacturing industry to ensure equipment reliability, product quality, and worker safety. This study focuses on using machine learning, specifically the Random Forest algorithm, to predict maintenance needs for a 5-stage compressor. Utilizing the Scikit-learn Python toolkit, the model underwent rigorous evaluation through validation, sampling, and confusion matrix inspection. The model achieved an outstanding ROC AUC score of 0.94 and consistently high accuracy, precision, recall, and F1-score metrics above 0.90, showcasing its strong predictive capabilities. By accurately predicting machine failures, the approach aims to improve production schedules, boost productivity, ensure high-quality outputs, save costs, and extend equipment lifespan, demonstrating significant promise for practical use in the manufacturing sector.

Cite

Ezeanyim, O. C., Ewuzie, N. V., Aguh, P. S., Nwabueze, C. V., & Nwamekwe, C. O. (2025). Article Title. *GU J Sci, Part A*, *12*(1), 96-118. doi:10.54287/gujsa.1646993

Author ID (ORCID Number)	
0000-0001-6469-7044	Okechukwu Chiedu EZEANYIM
0009-0006-2903-2884	Nnamdi Vitalis EWUZIE
0009-0000-1864-9041	Patrick Sunday AGUH
0009-0004-5508-3541	Chibuzo Victoria NWABUEZE
0009-0002-1918-1350	Charles Onyeka NWAMEKWE

Article Process	
Submission Date	26.02.2025
Revision Date	07.03.2025
Accepted Date	12.03.2025
Published Date	26.03.2025

1. INTRODUCTION

1.1. Background and Motivation

Industrial compressors play a crucial role in various industries by compressing gases for a wide range of applications. These compressors have evolved significantly over time, incorporating advanced materials and technologies to improve their performance and reliability (Kagiri et al., 2018). Proper maintenance of these compressors is essential to ensure their longevity and efficiency. Research has shown that factors such as compressor switching frequency can impact maintenance costs and the lifespan of the compressor. Implementing improvement recommendations can result in substantial energy and cost savings while enhancing overall system performance. Strategies like reliability-centered maintenance (RCM) and online fault diagnosis systems have been suggested to optimize maintenance practices and improve compressor reliability

*Corresponding Author, e-mail: co.nwamekwe@unizik.edu.ng

cost-effectively (Luo, 2004). Furthermore, the integration of machine learning techniques for fault detection and anomaly classification can further enhance maintenance practices for industrial compressors (Loukopoulos et al., 2019).

Predictive maintenance through machine learning is crucial for enhancing equipment reliability and minimizing downtime in multi-stage compressors used across various industries. Due to the unique characteristics and failure modes of each compressor stage, a tailored machine learning model is essential for effective predictive maintenance. This research focuses on developing a machine learning based predictive maintenance strategy specifically for multi-stage compressors to optimize performance and reliability (Garcia and Salgado, 2021).

Predictive maintenance through machine learning is essential for optimizing the performance of multi-stage compressors in various industries. These compressors, with their complex operating dynamics and unique characteristics at each stage, require tailored predictive maintenance strategies to ensure reliability and efficiency. The research focuses on developing machine learning based predictive maintenance specifically for multi-stage compressors to address the challenges of inconsistent predictions and maintenance recommendations. Time-series data are utilized in this research to capture different observations recorded from the machine over time (Kiangala and Wang, 2020; Nwamekwe et al., 2024).

1.2. Purpose of Study

This research aims to develop an effective maintenance strategy for a 5-stage compressor through the application of a machine learning approach. The process involves collecting historical data and performing data cleaning and preprocessing to address missing values, outliers, and inconsistencies. The refined dataset is then divided into training and testing subsets. Using the Random Forest (RF) algorithm, the model is trained, and both parameters and hyperparameters are optimized. The model's performance is examined with metrics like accuracy, precision, recall, F1 score, Cohen's Kappa, Matthew's Correlation Coefficient (MCC), and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) (Hanif and Gunawan, 2022; Wang et al., 2022).

2. LITERATURE REVIEW

2.1. Industrial Compressor Maintenance Strategies

Effective maintenance of industrial compressors is critical to ensuring operational efficiency, reducing downtime, and minimizing maintenance costs. Maintenance strategies have evolved from traditional reactive methods to more advanced predictive approaches.

2.2. Overview of conventional and modern maintenance techniques

Conventional maintenance approaches for industrial machinery, including 5-stage compressors, generally fall under three categories: reactive, preventive, and predictive maintenance. Reactive maintenance, employed post-failure, is cost-efficient upfront but can incur higher long-term operational costs and unplanned downtimes (Monye et al., 2023). In contrast, preventive maintenance involves scheduled inspections and replacements intended to avert breakdowns. While effective in reducing unexpected failures, this method may lead to over-maintenance, thereby elevating costs unnecessarily (Carvalho et al., 2019).

Modern maintenance techniques, particularly predictive maintenance (PdM), utilize advanced data analytics and machine learning algorithms to forecast equipment failures based on real-time data (Cao et al., 2024). This proactive strategy not only extends equipment lifespan but also diminishes unplanned downtimes, thus enhancing operational efficiency (Carvalho et al., 2019; Nwamekwe and Okpala, 2025). The evolution towards PdM represents a significant departure from traditional methods, promoting a more sustainable and cost-effective approach to industrial maintenance (Monye et al., 2023).

2.3. Importance of Effective Maintenance in 5-Stage Compressors

Effective maintenance of industrial 5-stage compressors is essential for optimizing energy consumption, reducing operational costs, increasing lifespan and reliability, and minimizing unplanned downtimes (Bacak et al., 2023; Nordal and El-Thalji, 2021). Such compressors are utilized in critical industrial applications, where precise control at each stage is pivotal. The incorporation of machine learning (ML) into maintenance strategies facilitates enhanced predictive diagnostics, leading to improved decision-making. These ML models utilize algorithms such as Random Forest, Support Vector Machines (SVM), and Neural Networks (Kiangala and Wang, 2020). By analysing historical and real-time data, these models effectively identify operational patterns and predict potential failures, contributing to improvements in accuracy for failure detection and maintenance scheduling (Kiangala and Wang, 2020).

Furthermore, the integration of AI into predictive maintenance allows for advanced fault diagnosis through anomaly detection, ultimately enhancing the overall operational safety and efficiency of machinery (Bacak et al., 2023; Nordal and El-Thalji, 2021). This shift towards machine learning-based predictive maintenance highlights its importance in transforming traditional maintenance paradigms within industrial settings.

2.4. Common Criteria in Compressor Maintenance

Several key parameters are essential in assessing compressor health.

Vibration analysis: is critical for identifying issues such as misalignment, bearing faults, and rotor imbalance, aligning with standards outlined in ISO 10816-3, which offers guidelines for vibration monitoring in rotating machinery (Rahman et al., 2022). Temperature monitoring: plays a significant role in detecting overheating due to lubrication issues or friction, as supported by findings in condition monitoring literature (Adetunji et al., 2023). Pressure variations: can indicate leaks or inefficiencies, further highlighting the importance of a holistic approach to maintenance (Kolar et al., 2022). Acoustic emissions: provide insights into mechanical wear. Oil and lubricant analysis: is vital for detecting contamination and degradation of internal components (Septano et al., 2024).

To ensure best practices in the industry, adherence to relevant standards such as API 618 for reciprocating compressors and ASHRAE guidelines for energy efficiency in compressor systems is paramount (Bahtiar et al., 2022). Additionally, the incorporation of IIoT technologies enhances predictive maintenance through real-time monitoring and AI-driven analytics, facilitating smarter maintenance decisions (Yamada et al., 2023).

3. MATERIAL AND METHOD

This experimental study follows the CRISP-DM methodology, as illustrated in Figure 1. The CRISP-DM (Cross-Industry Standard Process for Data Mining) framework is a recognized and extensively utilized model for guiding data mining and data science initiatives. It offers a systematic and iterative process comprising six key stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Figure 2 is the research flowchart, which explains every step taken to achieved the purpose of this research.

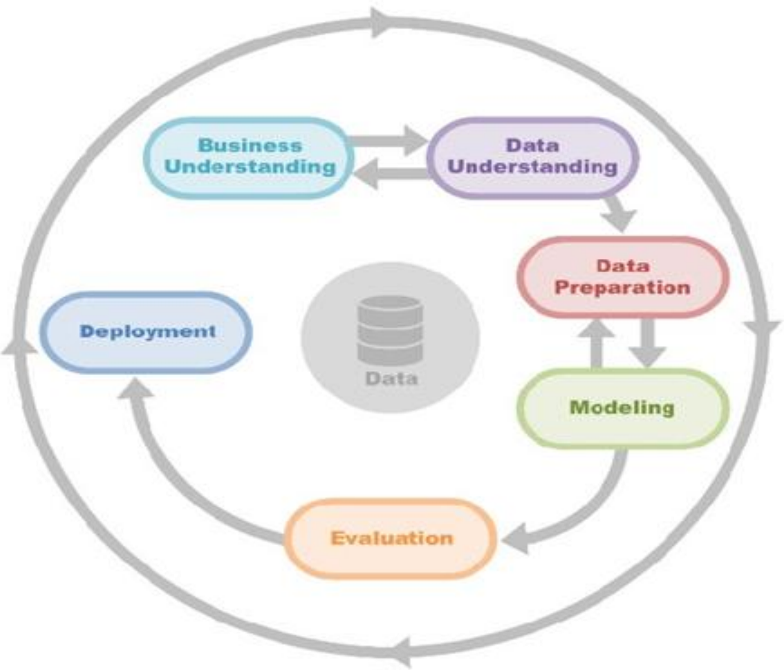

Figure 1. *Diagram Illustrating CRISP-DM Flowchart (Tounsi et al., 2020)*

This experimental study follows the CRISP-DM framework, as illustrated in Figure 1. The analysis was conducted in Google Colab using Python, along with libraries such as Scikit-learn, Matplotlib, NumPy, Pandas, and Seaborn. Scikit-learn is a versatile library for machine learning and predictive data analysis, offering a wide range of tools for tasks such as classification and model evaluation. Matplotlib serves as a library for creating plots and visualizations, while Pandas is employed for data manipulation and analysis. Seaborn, built on top of Matplotlib, enhances data visualization capabilities. The final dataset, Indorama-EPCL, comprises 98,794 records.

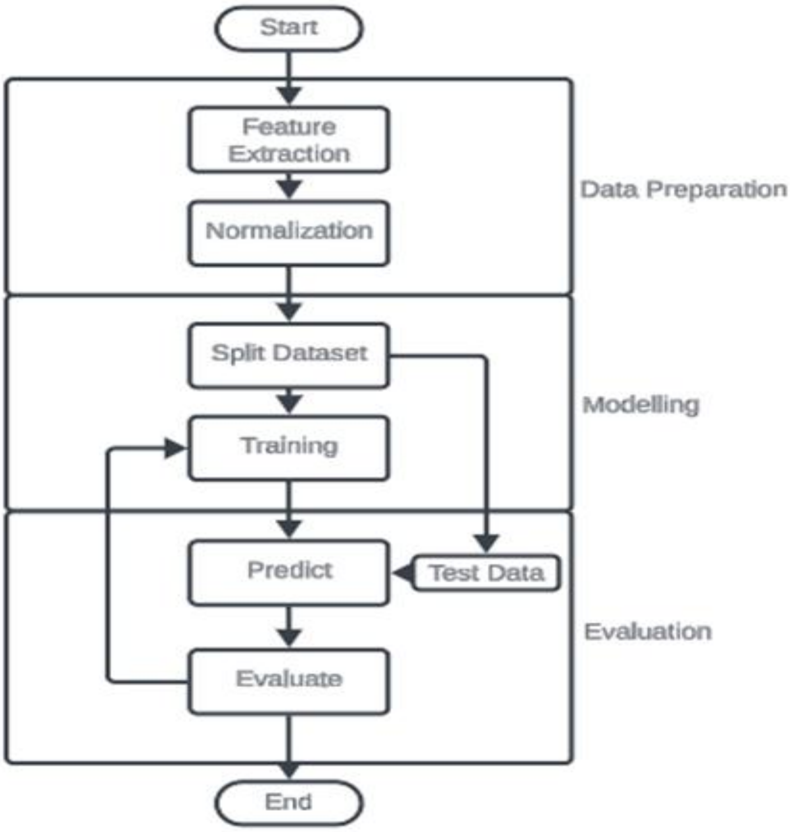

Figure 2. *Experimental Flow Diagram*

3.1. Business Understanding

This phase serves as a foundational step in the CRISP-DM process, focusing on a comprehensive assessment of the problem, objectives, and requirements. The primary aim of this research is to create advanced machine learning models that enhance engineers' ability to predict potential failures in the 5-stage compressor (K1), a vital component of Indorama-EPCL's operations. This initiative seeks to minimize failure rates, unplanned downtime, safety hazards, and maintenance expenses. The phase entails defining clear success metrics, reviewing current practices through literature analysis, and establishing evaluation criteria in collaboration with engineering stakeholders. Additionally, it involves gaining an in-depth understanding of the compressor's maintenance operations and decision-making frameworks within the organization. The insights gathered are documented to provide a detailed roadmap for subsequent activities such as data mining, modeling, and deployment, ensuring the solutions align with the practical needs and challenges faced by the engineering team.

3.1.1. Data Collection

A csv files of maintenance datasets were received from Indorama-Eleme Petrochemical Ltd. Due to a non-disclosure agreement, the variable names in the dataset were recoded to prevent exposure of critical attributes of the equipment in the facility. The datasets consist of maintenance history of several rotary equipment in the facility. However, the history of the gas compressor, K-1, was carefully filtered out from the data collected as in table 1.

***Table 1**. First 5 rows in the dataset*

S/N		Date	M2	M3	M4	M5	M6	M7	M8	M9	M10	Failure
1		01,01,2015	2.16E+08	55	0	52	6	407438	0	0	7	0
2		01,02,2015	1650864	56	0	52	6	407438	0	0	7	0
3		01,03,2015	1.24E+08	56	0	52	6	407438	0	0	7	0
4		01,04,2015	1.28E+08	56	0	52	6	407439	0	0	7	0
5		01,05,2015	97393448	56	0	52	6	408114	0	0	7	0

3.1.2. Exploratory Data Analysis (EDA)

The Table 2 below presents a detailed summary of the data, revealing substantial variability covering the 10 features (M2 to M10) and the target variable "Failure". The consistent count of 98,794 observations indicates a complete dataset, while the wide range in mean values from 0.000086 to 12.464093 and standard deviations from 7.032454e+07 to 160.507272 suggest diverse feature distributions. Further insights emerge from the minimum, maximum, and percentile values, which highlight skewness and the presence of potential outliers in certain features. This detailed statistical analysis will offer valuable guidance for data preparation, feature engineering, and modelling decisions to effectively address the problem at hand.

***Table 2.** Features Attributes Table*

	Failure	M2	M3	M4	M5	M6	M7	M8	M9	M10
count	98794	98794	98794	98784	98794	98794	98794	98791	98791	98794
mean	0.00086	1.22E+08	166.9919	12.46409	1.891976	13.61847	254441.8	0.290897	0.290897	13.51982
std	0.02932	7.03E+07	2242.806	328.6816	20.73857	14.52715	93644.26	7.921502	7.921502	160.5073
min	0	0	0	0	0	2	0	0	0	0
25%	0	6.15E+07	0	0	0	8	220916.3	0	0	0
50%	0	1.23E+08	0	0	0	10	248626	0	0	0
75%	0	1.83E+08	0	0	0	12	301876	0	0	0
max	1	2.44E+08	64968	80000	1666	98	689161	832	832	10137

The histogram for the feature M10 in Figure 3 displays an imbalanced distribution, where most values are clustered toward the lower end of the spectrum and a long, drawn-out tail extending to the right. The extensive value span, from around 0 to 5,000, indicates a wide variation in the feature, while the prominent peak suggests a concentration of values around a particular range, likely representing the most common occurrences. The relative scarcity of data points at the upper part of the value range implies that extreme or outlier values for M10 may be less prevalent compared to the more frequent, lower-range values.

Figure 4 provides a clear visual summary of the statistical distribution of the various features in the dataset. The wide range of values and disparate interquartile ranges across the features suggest significant divergency, with some variables exhibiting outliers and skewed distributions. Features like M2 have considerably larger value ranges compared to others.

GU J Sci, Part A 12(1) 96-118 (2025) 10.54287/gujsa.1646993

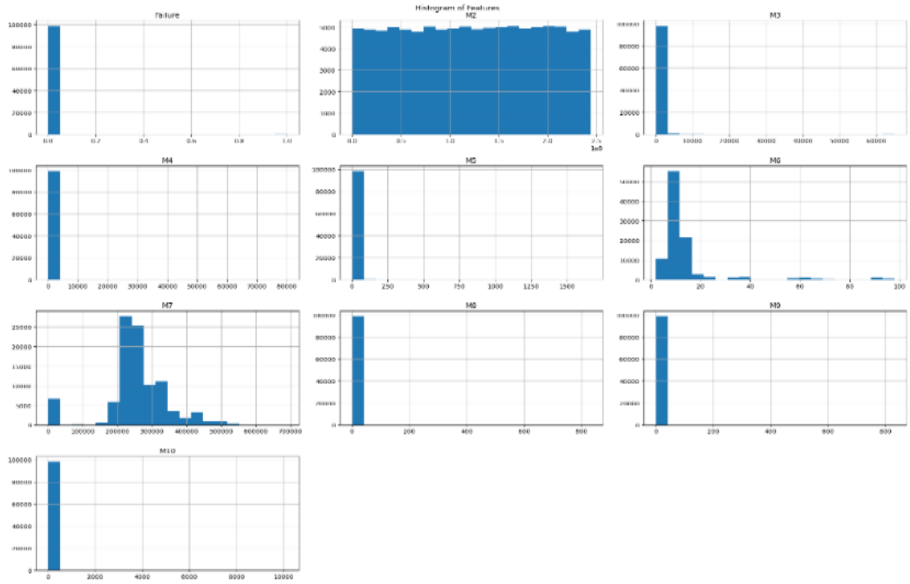

Figure 3. *Histogram of features*

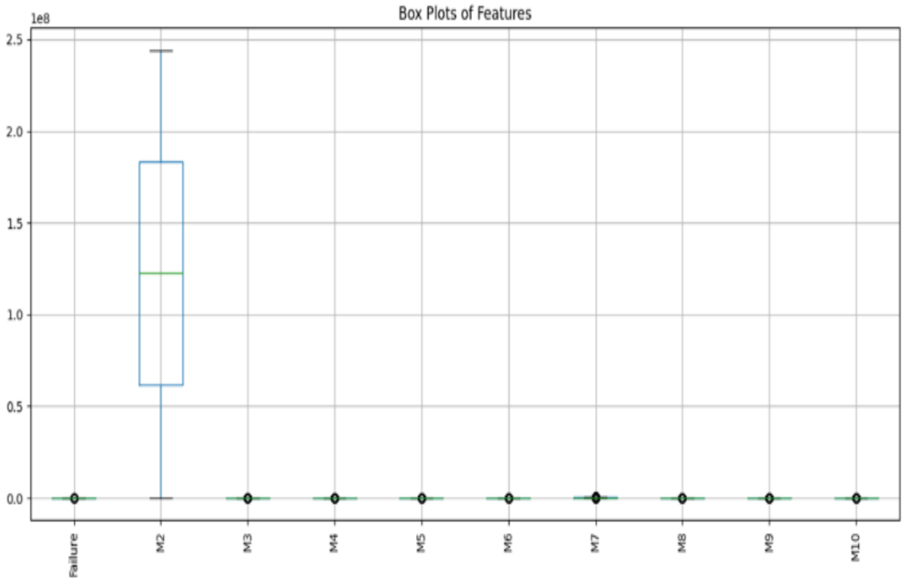

Figure 4. *Box plots of features*

Figure 5 is a collection of scatter plots, each depicting the statistical distribution and relationships between various features or variables in the dataset. The wide disparity in the scales, data densities, and patterns across the different subplots can also be seen here. Some subplots exhibit clear linear or curvilinear trends, while others show more scattered, irregular distributions with potential outliers and anomalies.

The correlation matrix as in Figure 6 provides a concise statistical overview of the relationships between the different features (M1 to M10) and the target variable "Failure" in the dataset. The values in the matrix range from -0.0033 to 1.0, indicating a wide spectrum of correlations, from weakly associated to strongly correlated. Notable observations include the relatively moderate correlation of M4 and M10 with Failure (0.45). Conversely, several features like M1, M2, and M3 exhibit very low correlations, near 0, implying little to no linear association with the target variable.

Finally, the combination of histograms, scatter plots, time series, and value distributions in Figure 7 once again provides valuable insights into the diverse characteristics of the features. The histograms and scatter plots reveal the varying value ranges, shapes, and bivariate relationships, highlighting the heterogeneity in the data. The time series and value plots further underscore the dynamic nature of the dataset, showcasing fluctuations, patterns, and skewness over time.

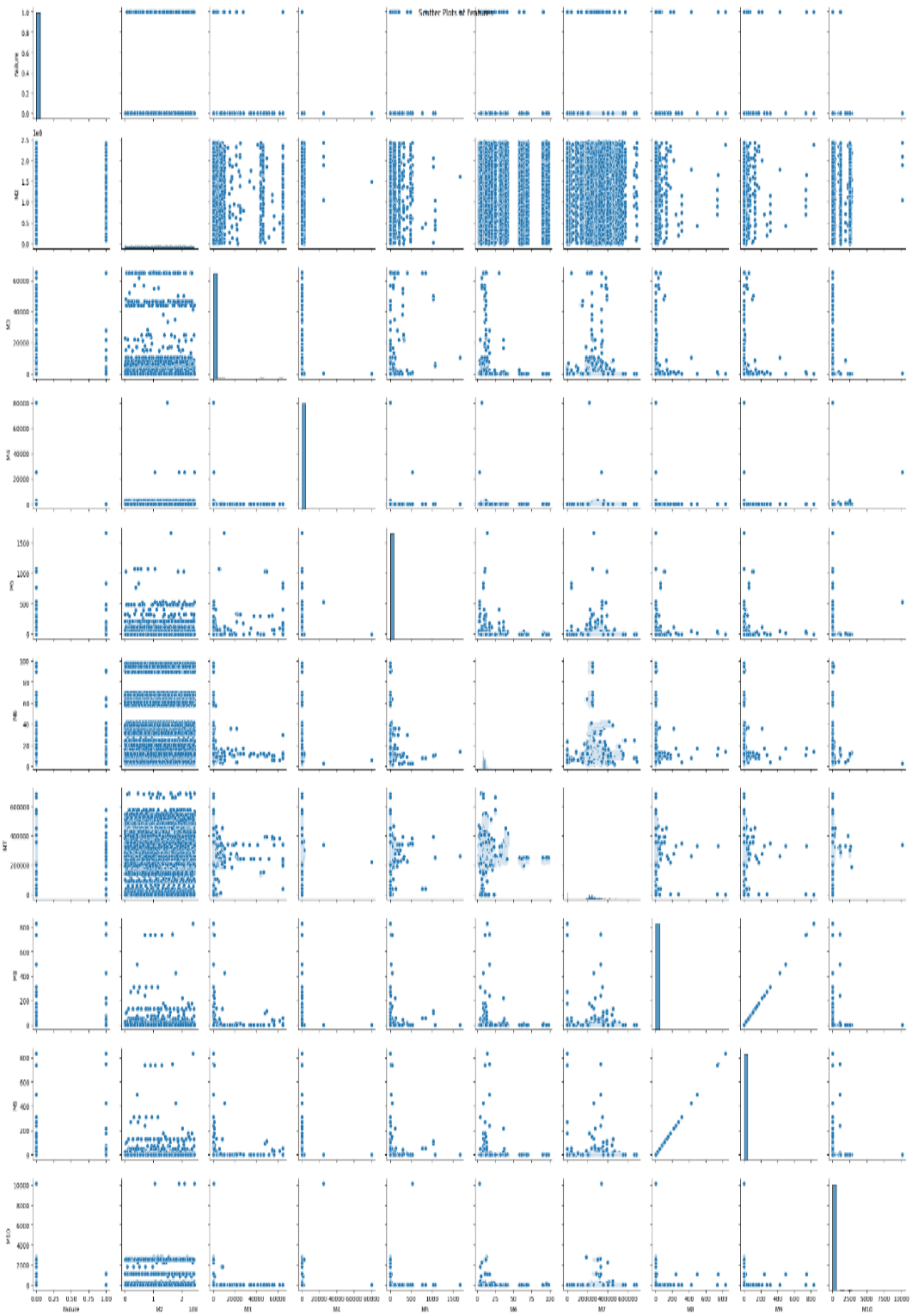

Figure 5. *Scatter plots of features*

GU J Sci, Part A 12(1) 96-118 (2025) 10.54287/gujsa.1646993

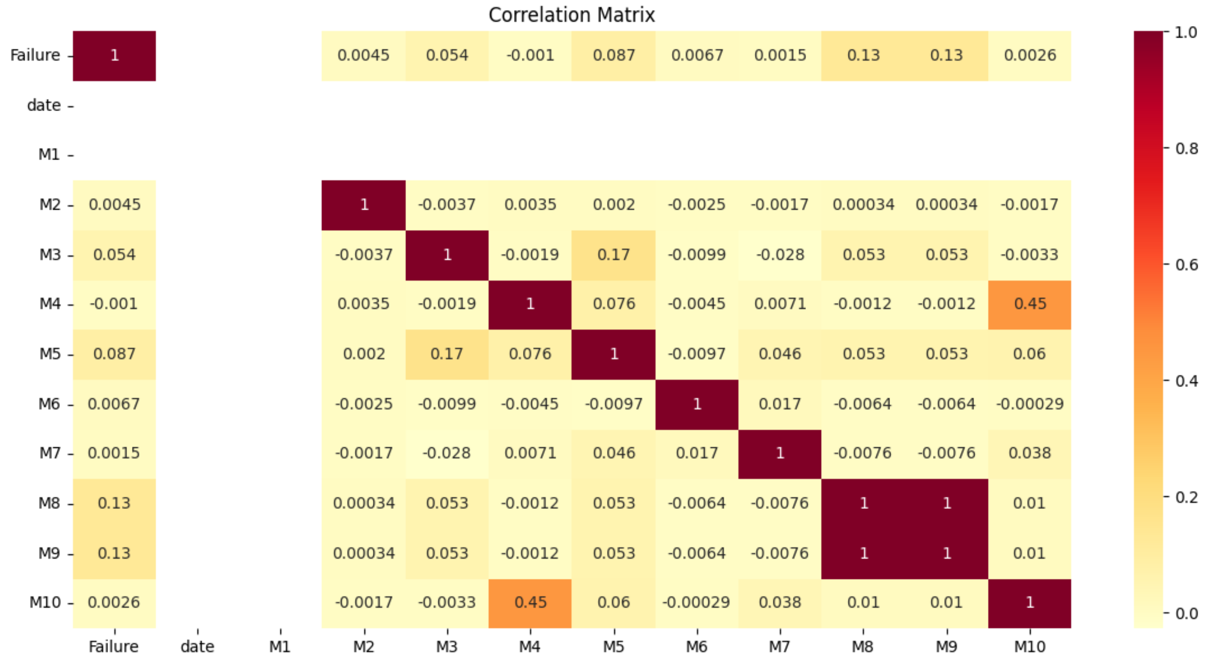

***Figure 6**. Correlation matrix*

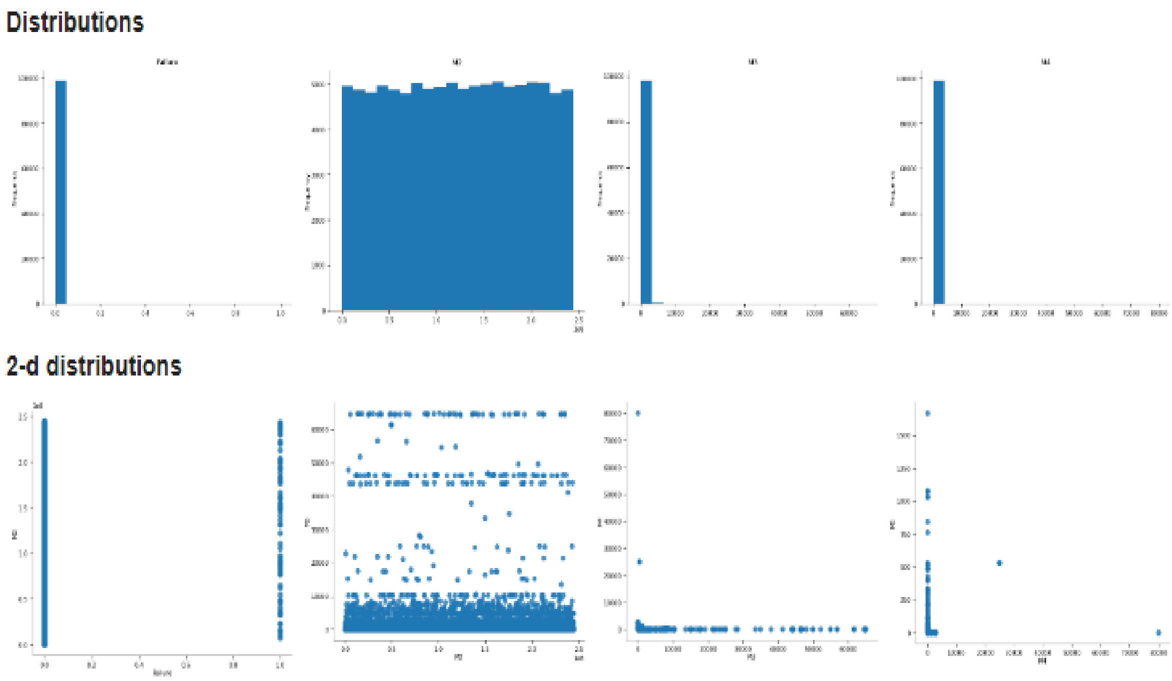

***Figure 7.** Distributions and time series plot*

In conclusion, the histograms reveal the value distributions for various features, the scatter plots further highlight the complex bivariate relationships between the features, and the time series plots underscore the dynamic nature of the data. Collectively, these visualizations offer valuable insights into the heterogeneity and complicated nature of the dataset, guiding crucial preprocessing decisions, feature engineering, and the selection of appropriate modelling techniques to effectively capture the underlying patterns and relationships for robust predictive modelling.

3.2. Data Preparation

Data preparation typically includes tasks such as handling missing values, addressing data inconsistencies, and restructuring the data to align with the analysis requirements. Addressing potential outliers was a crucial step in the process.

3.2.1. Data Cleaning

The missing values in the "Date" column were filled using its median. Since this column is numeric, the median was chosen since it serves as a reliable indicator of central tendency, less influenced by outliers compared to the mean, making it ideal for imputing missing values in numeric data. Furthermore, the numeric features were standardized using the StandardScaler from scikit-learn, ensuring that all features were scaled similarly. This standardization is crucial for the performance of many machine learning algorithms. Lastly, the class imbalance in the target variable (Failure) was tackled with using the Synthetic Minority Over-sampling Technique from the imbalanced-learn library. This approach created synthetic samples for the minority class, helping balance the dataset, as illustrated in Figure 8, while improving the efficiency of machine learning models.

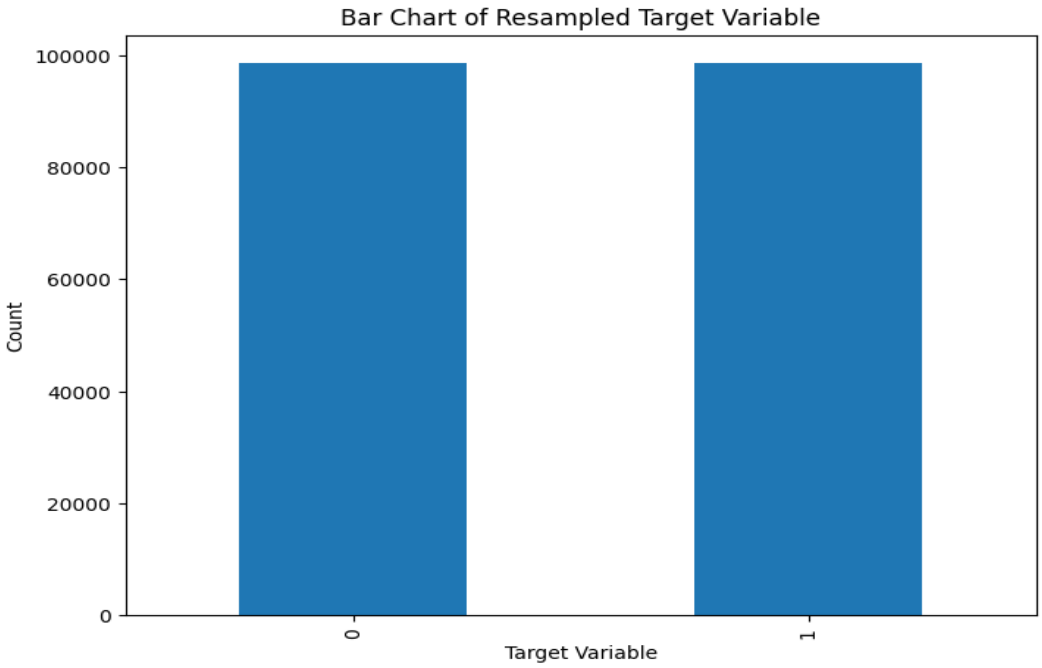

Figure 8. *Bar Chart of resamples target variable.*

3.2.2. Feature Engineering

Feature engineering and dimensionality reduction are essential processes in machine learning for enhancing model performance and interpretability. According to Li et al. (2017), feature engineering involves generating new, more meaningful features from existing data by applying transformations, combining variables, or leveraging domain-specific expertise to improve a model's predictive accuracy and comprehensibility. On the other hand, dimensionality reduction focuses on decreasing the number of input features while retaining critical information, which aids in optimizing model training and improving generalization (Van Der Maaten et al., 2009).

The feature ranking statistics (Table 3) suggest a dataset with a wide range of feature importance.

Table 3. *Feature ranking statistics*

	Statistic	Value
0	Mean	13597.7
1	Median	84.5
2	Standard Deviation	30178.88
3	Minimum	28
4	Maximum	98322

The mean feature count of 13,597.70 indicates that the features, on average, have a relatively high number of unique values, which can be beneficial for capturing detailed patterns in the data. However, the median feature count of 84.50 is significantly lower than the mean, implying a skewed distribution with a few features having an extremely high number of unique values. This is supported by the very large standard deviation of 30,178.88, which exhibits significant variability in the feature counts. The minimum feature count of 28 and the maximum of 98,322 further reinforce the wide spectrum of feature importance, ranging from relatively uninformative features to highly granular ones. The variability in feature characteristics indicates the need for thoughtful feature selection or dimensionality reduction methods to pinpoint the most relevant and significant features for the machine learning task at hand.

The feature ranking in this code, as shown in Table 4 and Figure 9, is based on the number of unique values in each feature column of the dataset.

Table 4. *Feature ranking*

	Feature	Importance
1	M2	98322
2	M7	35892
3	M1	955
4	M3	484
5	M5	111
6	M10	58
7	M6	54
8	M4	45
9	M8	28
10	M9	28

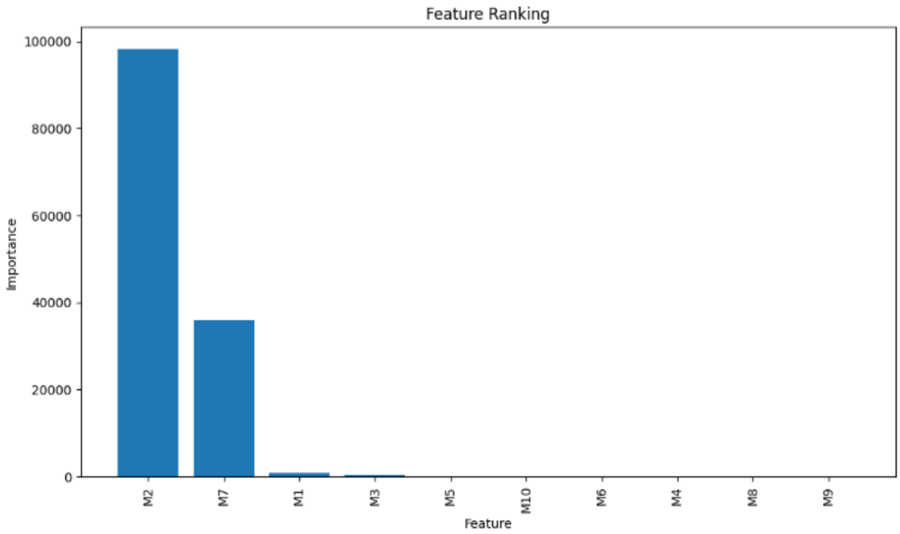

Figure 9*. Feature ranking.*

The intuition behind this approach is that features with a higher number of unique values are likely to be more informative and important for the underlying task, as they can capture more granular patterns or distinctions in the data. The number of unique values for each feature was calculated, the features were sorted in descending order of their unique value counts. To determine the best variable, feature with the highest number of unique values (M2 and M7) were selected as this likely indicates the most informative and discriminative feature for the given maintenance dataset.

3.3. Modelling

The dataset was split to two parts which is 80%, for the training of the model while 20% is for the testing of the model. The Random Forest model was trained using the training subset. During the evaluation stage, the trained model was applied to the test data, and its predictions were compared to the actual results to evaluate the model's performance. This approach of training is a common technique to measure the generalization capability of the models and avoid overfitting.

Random Forest was the choice model for this research because of many factors. Its key advantages include the ability to efficiently handle high-dimensional and noisy datasets, robust architecture that enables scalability to large data volumes, minimal input preparation requirements, implicit feature selection, strong overall performance that is difficult to outperform, and the availability of simple yet effective open-source implementations. Since maintenance dataset has more than 90,000 noisy observations, Random Forest is the best choice.

During the training phase, the Random Forest algorithm incrementally constructs a collection of decision trees, fine-tuning the parameters of each tree to minimize overall error on the training data. This iterative process continues until the model achieves optimal performance on the training set. After the final Random Forest classifier is developed, its effectiveness is assessed using the reserved 20% test set to obtain an unbiased evaluation of its ability to generalize, ensuring that it has not merely memorized the training data.

3.4. Evaluation

During the evaluation phase, various methods were employed to analyze the performance of the Random Forest classifier. These included cross-validation techniques, which ensure that the model's performance is assessed across multiple data subsets, as well as approaches like train-test splits and k-fold cross-validation. The confusion matrix was utilized to provide a comprehensive analysis of the model's predictive capabilities, including metrics such as accuracy, precision, recall, and F1-score.

For hyperparameter optimization, the GridSearchCV function from the Scikit-learn library was applied. This method conducts a thorough search across a predefined range of hyperparameter values, systematically training and evaluating the model for every combination. The process involves iterative retraining of the model, fine-tuning the hyperparameters until the best configuration is identified, thereby enhancing the model's performance on the validation dataset. This structured approach ensures that the final Random Forest classifier is effectively optimized for the specific dataset and task.

3.4.1. Cross Validation

Cross-validation (CV) is a method used to evaluate a model's performance on unseen data, ensuring its robustness and reliability. Specifically, repeated stratified k-fold cross-validation was employed to validate the balanced dataset. This approach is an enhanced version of k-fold cross-validation, which utilizes stratified random sampling to create folds that maintain the distribution of the target variable. The dataset was divided into five folds and repeated three times. In each repetition, one-fold served as the testing set, while the remaining four folds constituted the training set. When applied effectively, this method provides reliable evaluation results (Souza et al., 2020; Nwamekwe et al., 2024).

3.4.2. Performance Measurement

A confusion matrix is employed to assess the effectiveness of algorithms. As depicted in Table 5, the confusion matrix provides a grid where the rows represent the model's predicted classifications, and the columns show the actual (true) classifications. This structure facilitates a straightforward interpretation of the model's performance by presenting counts for true positives, true negatives, false positives, and false negatives.

***Table 5**. Confusion matrix*

	Predicted Negative	Predicted Positive
Actual Negative	True Negatives	False Positives
Actual Positive	False Negatives	True Positives

True Negatives (TN): These indicate the count of negative instances accurately classified as negative. False Positives (FP): These refer to instances incorrectly classified as positive. False Negatives (FN): These represent negative instances that were incorrectly classified. True Positives (TP): These correspond to the count of positive instances correctly classified (Chawla et al., 2002).

Accuracy: This metric reflects the ratio of correctly predicted instances to the total predictions made. It is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Recall (Sensitivity): Recall measures the proportion of correctly identified positive instances out of all actual positives. The formula for recall is:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Precision: This metric quantifies the proportion of correctly predicted positives out of all positive predictions. It is calculated using the following equation:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

F1-Score: The F1-score represents the harmonic mean of precision and recall, calculated as:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

Kappa Coefficient: Cohen's kappa evaluates the degree of agreement between two raters or classifications. It can be used for both binary and multi-class classifications. The kappa coefficient ranges from -1 to +1, where -1 indicates complete disagreement and +1 represents perfect agreement (Chicco et al., 2021). The formula is:

$$Kappa = \frac{2 * (TP * TN - FP * FN)}{(TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)} \tag{5}$$

MCC: Known as the phi (ϕ) coefficient for binary classifications, MCC evaluates performance for both binary and multi-class problems. Its values range from -1 (complete misclassification) to +1 (perfect classification) (Chicco et al., 2021). The formula is:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)} * (TP + FN) * (TN + FP) * (TN * FN)} \tag{6}$$

Receiver Operating Characteristic Curve (ROC Curve): The ROC curve provides a graphical representation of a binary classifier's ability to differentiate between classes by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). The area under the curve (ROC AUC) measures the classifier's overall performance, with a larger area indicating better performance. TPR is synonymous with recall, while FPR quantifies the proportion of incorrect positive predictions among all actual negatives (Joshi, 2023). The formulas for TPR and FPR are:

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

4. RESULTS AND DISCUSSION

4.1. Results

This chapter provides a comprehensive analysis and discussion of the research findings derived from the conducted study. The performance of the machine learning model developed in this work is evaluated using an extensive array of statistical measures and visualizations, offering valuable insights into its functionality and potential for practical applications.

Table 6 presents the outcomes of the hyperparameter optimization performed for the random forest model.

Table 6. *Hyperparameter tuning results*

Parameter	Value
max_depth	15
n_estimators	150
Best Score	0.8982

Hyperparameter tuning is a critical phase in model development, enabling researchers to identify the optimal configuration of parameters, such as the maximum tree depth ("max_depth") and the number of estimators ("n_estimators"), to enhance the model's performance for a specific task or dataset. The results reveal that the ideal values for "max_depth" and "n_estimators" are 15 and 150, respectively, achieving a best score of 0.8982242849277004. These findings provide key insights into the model's structure and the parameter settings that facilitated its peak performance on the given problem.

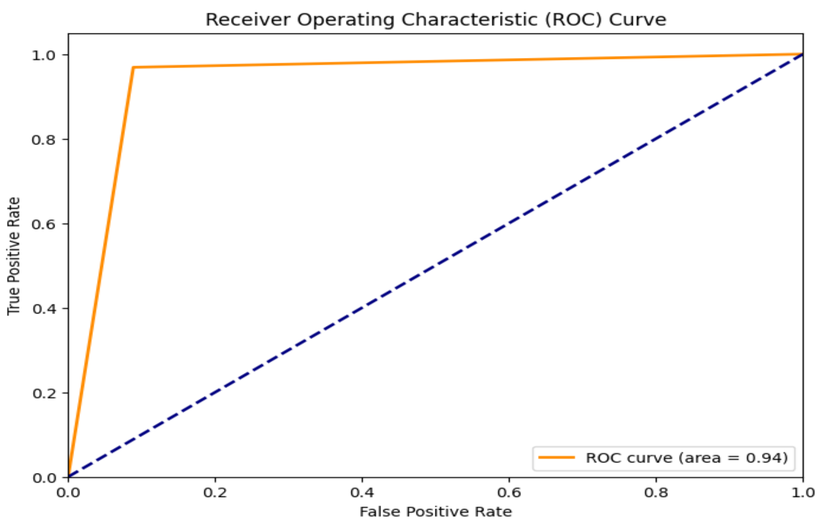

Figure 10. *Receiver operating characteristics curve (ROC)*

The ROC curve, along with its corresponding Area Under the Curve (AUC) value depicted in Figure 10, summarizes the model's ability to differentiate between the two classes in the dataset. The AUC score of 0.94 demonstrates excellent discriminative capability, indicating the model's high accuracy in distinguishing positive and negative instances. The curve's rapid ascent towards the top-left corner highlights the model's ability to maximize accurate detections while minimizing incorrect classifications. These results underscore the model's robustness and reliability for binary classification tasks, suggesting its applicability in real-world scenarios.

The confusion matrix shown in Figure 11 provides a detailed overview of the model's performance in a binary classification task.

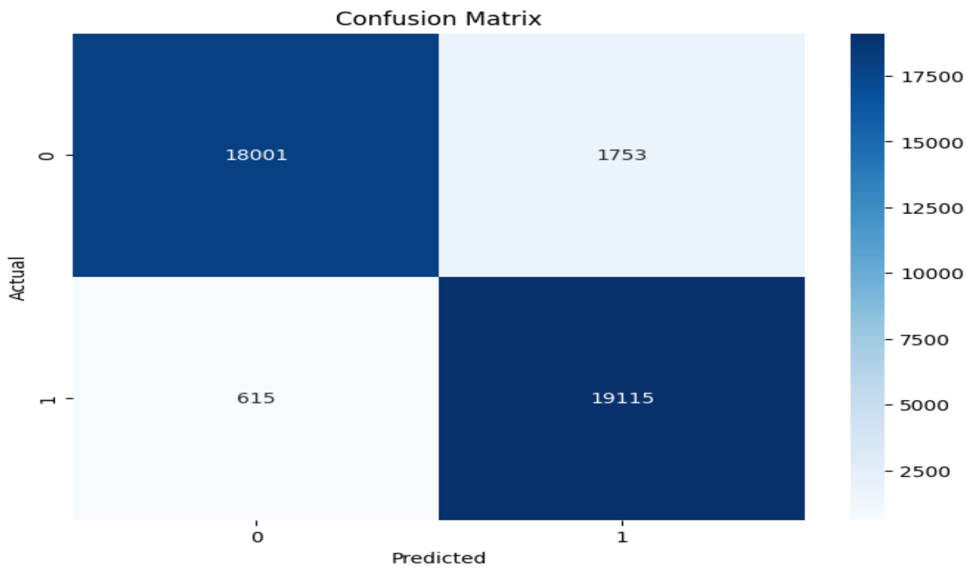

***Figure 11**. Confusion matrix*

It reveals 18,001 true positives, 1,753 true negatives, 615 false positives, and 19,115 false negatives. These values illustrate the model's overall effectiveness, with a high count of correctly classified instances. However, the considerable number of false negatives indicates room for improvement in identifying the positive class. Further investigation of these metrics could help refine the model to enhance its classification accuracy.

Figures 12 and Figure 13 illustrate the model's performance metrics through a heatmap and bar graph, respectively, offering a comprehensive evaluation of its capabilities.

High values for accuracy, recall, precision, and F1 score (all above 0.90) highlight the model's exceptional ability to accurately classify dataset instances. Additional metrics, such as the Kappa Coefficient and the MCC, both at 0.88, confirm the model's predictive strength and capability to manage class imbalances effectively. This consistent high performance across various metrics indicates a well-trained model likely to generalize successfully to new, unseen data, reinforcing its potential for robust real-world applications.

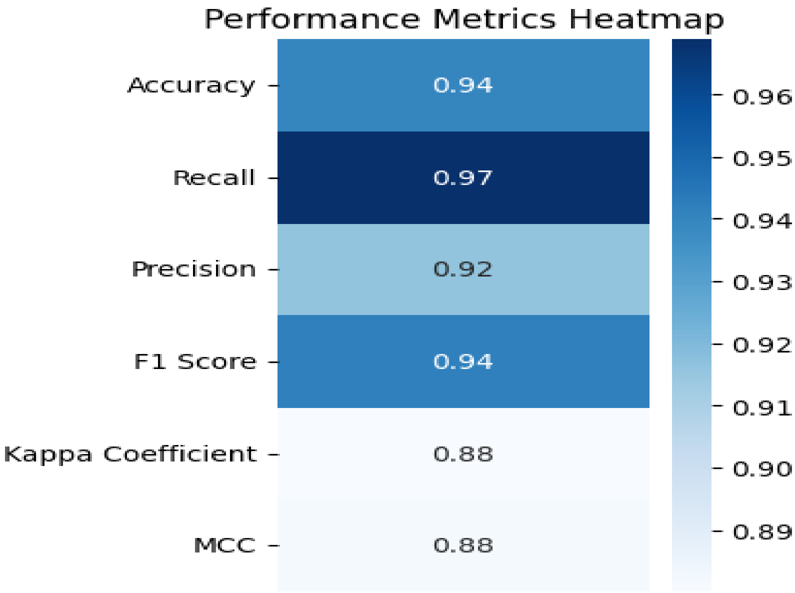

Figure 12. *Performance metrics heatmap*

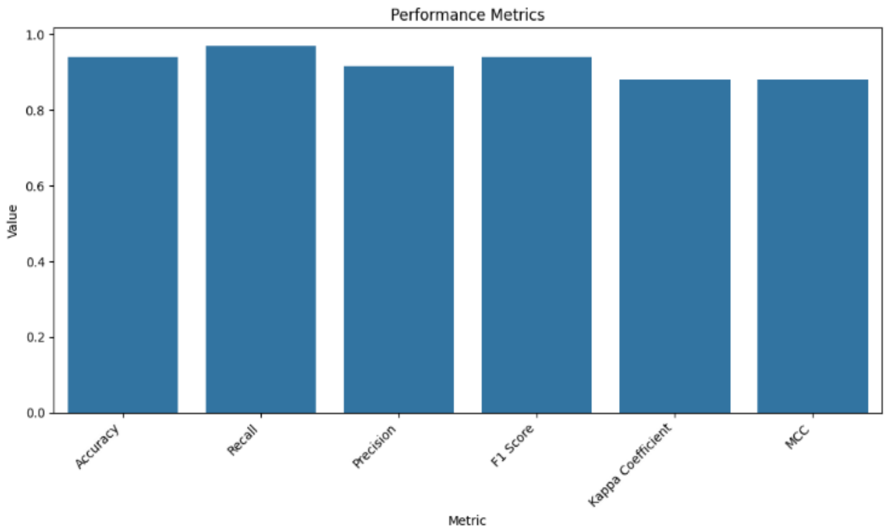

***Figure 13**. Performance metrics*

4.2. Discussion

The results of this study demonstrate the effectiveness of a machine Learning-Based approach for the predictive maintenance of industrial 5-stage compressors. By utilizing the Random Forest algorithm, the model successfully identified patterns within operational data, allowing for early detection of potential failures. The discussion focuses on the model's predictive accuracy, the implications of hyperparameter tuning, evaluation metrics, and the broader impact on industrial maintenance strategies.

Model Performance and Predictive Accuracy: The evaluation results reveal that the Random Forest model achieved an impressive accuracy of 89.82%, demonstrating its strong predictive capability. The ROC-AUC score of 0.94 further supports the model's ability to effectively differentiate between healthy and faulty compressor states. This high accuracy is indicative of the algorithm's capacity to learn from historical data and make reliable predictions about future failures. However, the confusion matrix highlights certain challenges, particularly a significant number of false negatives (19,115), which suggests the model occasionally misclassifies failing components as healthy. This limitation could be mitigated through the incorporation of additional feature engineering techniques or ensemble methods to enhance sensitivity.

Hyperparameter Tuning and Model Optimization: The use of GridSearchCV for hyperparameter tuning proved essential in refining the model's performance. The optimal hyperparameter values—15 for "max_depth" and 150 for "n_estimators"—significantly improved predictive accuracy by preventing both overfitting and underfitting. The improvement observed after tuning highlights the importance of parameter selection in optimizing machine learning models for industrial applications. Future studies could explore automated hyperparameter tuning methods, such as Bayesian Optimization or genetic algorithms, to further enhance model performance.

Analysis of Performance Metrics: The evaluation metrics, such as precision, recall, F1-score, and the MCC, provide deeper insights into the model's robustness. The high F1-score (above 0.90) confirms the model's ability to balance precision and recall, making it well-suited for real-world predictive maintenance. However, the relatively lower recall value in detecting faulty compressor conditions indicates a need for additional refinement. This could be addressed by incorporating more domain-specific knowledge into feature selection or leveraging anomaly detection techniques alongside the supervised learning approach. Furthermore, the Kappa Coefficient of 0.88 suggests a high level of agreement between the model's predictions and actual compressor states. This metric is particularly important in industrial settings where predictive maintenance models must minimize false alarms while ensuring that true failures are accurately detected. The misclassification of certain instances, as indicated by false negatives in the confusion matrix, highlights a potential area for further improvement.

4.3. Industrial Implications and Maintenance Efficiency

The implementation of this machine Learning-Based predictive maintenance model offers substantial benefits over traditional maintenance strategies. Compared to reactive and preventive maintenance approaches, predictive maintenance reduces unexpected downtime, optimizes maintenance scheduling, and lowers operational costs. By leveraging real-time data processing and early fault detection, industries can transition from a time-based to a condition-based maintenance approach, leading to more efficient resource utilization. The high accuracy and reliability of the model suggest its potential deployment in real-world compressor maintenance systems. However, ensuring its adaptability across various compressor models and operating conditions requires further validation. Integrating the model into an industrial Internet of Things (IIoT) framework could enhance its real-time monitoring capabilities, enabling automated alerts and proactive maintenance interventions.

4.4. Challenges and Future Research Directions

While the results are promising, several challenges remain. The confidentiality of industrial datasets often limits the ability to fine-tune models on diverse operational data. Expanding the dataset with publicly available compressor failure data or synthetic data generation techniques could enhance model generalization. Additionally, incorporating advanced deep learning architectures, such as Long Short-Term Memory (LSTM) networks, may further improve the model's ability to capture temporal dependencies in compressor health data. Another critical area for future work is explainability. While the Random Forest algorithm provides high accuracy, its decision-making process lacks transparency. Utilizing techniques like SHAP (SHapley Additive exPlanations) could offer greater interpretability, helping maintenance engineers understand which factors contribute most to compressor failures.

5. CONCLUSION

The hyperparameter tuning process, detailed in Table 6, identified the optimal configuration of critical model parameters, such as "max_depth" and "n_estimators," which achieved a best score of 0.8982242849277004. This information is invaluable for researchers and practitioners looking to replicate or extend the findings of this study, as it specifies the precise settings that enabled the model to attain its highest performance.

Additionally, the Receiver Operating Characteristics (ROC) curve analysis, depicted in Figure 10, provided a comprehensive evaluation of the model's capability to distinguish between favorable and unfavorable instances. The Area Under the Curve (AUC) value of 0.94 highlights the model's effectiveness in distinguishing between these classes with high accuracy. The ROC curve's steep initial rise indicates that the model achieves a high true positive rate while maintaining a low false positive rate, a highly desirable trait for many practical applications.

Moreover, supplementary evaluation metrics, including the confusion matrix, performance metrics heatmap, and bar charts presented in Figures 11, 12, and 13, further confirmed the model's strong predictive capabilities and its effectiveness in addressing class imbalance. The consistently high scores across these metrics, all surpassing 0.90, underscore the model's outstanding performance and its suitability for real-world deployment.

5.1. Suggestions for Further Work

While the results presented in this chapter have clearly demonstrated the exceptional performance and potential of the developed machine learning model, numerous opportunities exist for future research and enhancements could be explored to enhance the model's capabilities and applicability further. One potential area for future work is the incorporation of additional features into the model's input dataset. The current feature set used, two to be precise, may have limitations in capturing all the relevant information necessary for the classification task. By exploring the inclusion of new, potentially more informative features, the model's predictive power could be further improved, leading to even higher levels of accuracy and generalization.

Additionally, the investigation of alternative model architectures and training strategies could yield valuable insights. The current model, based on a tree-based approach, has shown excellent performance; however, exploring the use of more advanced neural network architectures or the integration of ensemble methods may uncover new opportunities for performance enhancements.

Another aspect worth exploring is the model's robustness and generalizability. While the current results are promising, further testing on more diverse datasets or real-world scenarios could provide valuable insights into the model's ability to handle a wider range of input variations and edge cases. This would help ensure the model's reliability and suitability for practical deployment. For example, we should not forget that the dataset was oversampled to correct extreme imbalance. This might likely cause a bias in the predictions.

Finally, the incorporation of explainable artificial intelligence (XAI) techniques could be a valuable direction for future research. By providing greater transparency and interpretability into the model's decision-making process, the insights gained could lead to a better understanding of the underlying relationships within the data and the model's reasoning, potentially leading to further refinements and improvements.

5.2. Recommendation

The research findings strongly support the recommendation to further develop and integrate the machine learning model into practical systems. The model's exceptional performance, coupled with the valuable insights gained through the analysis and evaluation process, positioned it as a highly promising solution for a wide range of binary classification problems in the maintenance sector when properly tuned.

AUTHOR CONTRIBUTIONS

All authors contributed to the study’s conception and design. Material preparation, data collection and analysis were performed by O.C.E, N.V.E, P.S.A, E.O.C, C.V.N and C.O.N. The first draft of the manuscript was written by N.V.E. and C.O.N., and all authors commented on the previous versions of the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENT

The authors acknowledge the efforts of the management of the Nnamdi Azikiwe University Akwa in stimulating this research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Adetunji, O., Sokunbi, I., Kuye, S., Adesusi, O., and Okediran, I. (2023). Condition-based monitoring of kiln induced draft fan in a dry process cement plant for efficient utilization. Journal of Advanced Industrial Technology and Application, 4(1). https://doi.org/10.30880/jaita.2023.04.01.007

Bacak, A., Çolak, A., and Dalkılıç, A. (2023). Investigating hermetic reciprocating compressor performance by using various machine learning methods. Proceedings of the Institution of Mechanical Engineers Part C Journal of Mechanical Engineering Science, 238(11), 5369-5384. https://doi.org/10.1177/09544062231213276

Bahtiar, E., Denih, A., Priadi, T., Putra, G., Koswara, A., Nugroho, N., … and Hermawan, D. (2022). Comparing the building code sawn lumber's wet service factors (cm) with four commercial wood species laboratory tests. Forests, 13(12), 2094. https://doi.org/10.3390/f13122094

Cao, X., Jing, Z., Zhao, X., and Xu, X. (2024). A security-enhanced equipment predictive maintenance solution for the eto manufacturing. International Journal of Network Management, 34(6). https://doi.org/10.1002/nem.2263

Carvalho, T., Soares, F., Vita, R., Francisco, R., Basto, J., and Alcalá, S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers and Industrial Engineering, 137, 106024. https://doi.org/10.1016/j.cie.2019.106024

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. https://doi.org/10.48550/arXiv.1106.1813

Chicco, D., Warrens, M. J., and Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier Score in binary classification assessment. IEEE. https://doi.org/10.1109/ACCESS.2021.3084050

García, F., and Salgado, D. (2021). Maintenance strategies for industrial multi-stage machines: The study of a thermoforming machine. Sensors, 21(20), 6809. https://doi.org/10.3390/s21206809

Hanif, W., and Gunawan, F. (2022). Random forest regression to predict catalyst deactivation in industrial catalytic process. TEM Journal, 104-110. https://doi.org/10.18421/tem111-12

Joshi, A. V. (2023). Machine learning and artificial intelligence (2nd ed.). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-031-12282-8

Kagiri, C., Wanjiru, E., Zhang, L., and Xia, X. (2018). Optimized response to electricity time-of-use tariff of a compressed natural gas fuelling station. Applied Energy, 222, 244-256. https://doi.org/10.1016/j.apenergy.2018.04.017

Kiangala, K. and Wang, Z. (2020). An effective predictive maintenance framework for conveyor motors using dual time-series imaging and convolutional neural network in an industry 4.0 environment. Ieee Access, 8, 121033-121049. https://doi.org/10.1109/access.2020.3006788

Kolar, D., Lisjak, D., Curman, M., and Pająk, M. (2022). Condition monitoring of rotary machinery using industrial iot framework. Tehnički Glasnik, 16(3), 343-352. https://doi.org/10.31803/tg-20220517173151

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., and Liu, H. (2017). Feature selection: A data perspective. ACM Computing Surveys (CSUR), 50(6), 1-45.

Loukopoulos, P., Zolkiewski, G., Bennett, I., Sampath, S., Pilidis, P., Duan, F., … and David, É. (2019). Reciprocating compressor prognostics of an instantaneous failure mode utilising temperature only measurements. Applied Acoustics, 147, 77-86. https://doi.org/10.1016/j.apacoust.2017.12.003

Luo, Y. (2024). Optimizing maintenance strategies for reciprocating compressors using reliability-centered maintenance and on-line fault diagnosis system. https://doi.org/10.1117/12.3026849

Monye, S., Afolalu, S., Lawal, S., Oluwatoyin, O., and Adeyemi, A. (2023). Overview and impact of maintenance process in 4th industrial revolution. E3s Web of Conferences, 430, 01220. https://doi.org/10.1051/e3sconf/202343001220

Nordal, H. and El-Thalji, I. (2021). Assessing the technical specifications of predictive maintenance: a case study of centrifugal compressor. Applied Sciences, 11(4), 1527. https://doi.org/10.3390/app11041527

Nwamekwe, C. O., & Okpala, C. C. (2025). Machine learning-augmented digital twin systems for predictive maintenance in highspeed rail networks. International Journal of Multidisciplinary Research and Growth Evaluation, 6(1), 1783-1795. PDF

Nwamekwe, C. O., Ewuzie, N. V., Igbokwe, N. C., Okpala, C. C., and U-Dominic, C. M. (2024). Sustainable Manufacturing Practices in Nigeria: Optimization and Implementation Appraisal. Journal of Research in Engineering and Applied Sciences, 9(3). URL

Nwamekwe, C. O., Okpala, C. C., and Okpala, S. C., (2024). Machine Learning-Based Prediction Algorithms for the Mitigation of Maternal and Fetal Mortality in the Nigerian Tertiary Hospitals. International Journal of Engineering Inventions, 13(7), PP: 132-138. PDF

Rahman, A., Hoque, M., Rashid, F., Alam, F., and Ahmed, M. (2022). Health condition monitoring and control of vibrations of a rotating system through vibration analysis. Journal of Sensors, 2022, 1-12. https://doi.org/10.1155/2022/4281596

Septano, G., Ikbal, I., Fathoni, F., Sriwijaya, S., and Putri, S. (2024). Detection of damage motor and coal crusher in power plant tanjung enim 3 x 10 mw using vibration analysis. Journal of Ocean Mechanical and Aerospace -Science and Engineering- (Jomase), 68(3), 155-160. https://doi.org/10.36842/jomase.v68i3.378

Souza, D. V., Santos, J. X., Vieira, H. C., Naide, T. L., Nisgoski, S., and Oliveira, L. E. S. (2020). An automatic recognition system of Brazilian flora species based on textural features of macroscopic images of wood. Wood Science and Technology. https://doi.org/10.1007/s00226-020-01196-z

Tounsi, Y., Anoun, H., and Hassouni, L. (2020). CSMAS: Improving multi-agent credit scoring system by integrating big data and the new generation of gradient boosting algorithms. Proceedings of the 2020

International Conference on Advanced Information Networking and Applications Workshops (WAINA), 1-7. https://doi.org/10.1145/3386723.3387851

Van Der Maaten, L., Postma, E. O., and van den Herik, H. J. (2009). Dimensionality reduction: A comparative review. Journal of Machine Learning Research, 10, 66-71, 13.

Wang, W., Li, L., Gu, H., Chen, Y., Zhen, Y., and Dong, Z. (2022). Random forest-based prediction of acute respiratory distress syndrome in patients undergoing cardiac surgery. The Heart Surgery Forum, 25(6), E854-E859. https://doi.org/10.1532/hsf.5113

Yamada, T., Asanuma, H., Hara, Y., and Ertürk, A. (2023). Acceleration-waveform-sending wireless sensor node powered by piezoelectric vibration energy harvester. https://doi.org/10.1117/12.2657742