Proceedings of the 6th Asia Pacific Conference on Industrial Engineering and Operations Management
Bali, Indonesia, August 5-7, 2025

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/AP06.20250115
Published: August 5, 2025

System of Predictive Maintenance in Palm Oil Mill

Nursyuhadah Saleh, Jafri Mohd Rohani, Nurul Rawaida Ain Burhani and Puteri Mawardati Yasir
Faculty of Mechanical Engineering
Universiti Teknologi Malaysia (UTM)
81300 Skudai, Johor Malaysia
nursyuhadahsaleh01@gmail.com, jafrimr@utm.my, nurulrawaidaain@utm.my, puterimawardati@gmail.com

Abstract

The boiler station is a critical section in palm oil mills as it supplies steam for sterilization and energy generation. However, a case study revealed that the boiler station operated with an Overall Equipment Effectiveness (OEE) of only 27.03%, far below the 85% world-class standard. The main contributors to this inefficiency were reactive maintenance practices, delayed failure responses, and aging components with no early detection mechanisms, resulting in frequent breakdowns and prolonged downtimes. This research was conducted with three objectives: (1) to evaluate and analyze the root causes of low OEE performance, (2) to develop a predictive maintenance system using logistic regression to forecast component failures based on historical maintenance data, and (3) to integrate an automated alert system that notifies engineers of high-risk components before failures occur. Logistic regression was selected for its simplicity, interpretability, and suitability for environments with limited digital infrastructure, making it ideal for palm oil mills with structured Excel-based records. The predictive system was developed using Python and Streamlit, enabling engineers to upload maintenance logs, receive failure probability predictions, and visualize high-risk components through an interactive web interface. The model achieved a prediction accuracy of 77%, demonstrating its effectiveness in shifting maintenance practices from reactive to predictive, improving equipment reliability, and aligning with Total Productive Maintenance (TPM) and Industry 4.0 principles.

Keywords
Overall Equipment Effectiveness, Total Productive Maintenance, Preventive Maintenance, Predictive Maintenance

1. Introduction
The boiler station in palm oil mills serves as a crucial unit that generates steam required for sterilization, electricity production, and other thermal operations. However, many palm oil mills continue to face persistent operational inefficiencies due to outdated maintenance strategies. Reactive maintenance where repairs are only initiated after a breakdown—remains a common practice. This leads to unplanned downtimes, inefficient resource allocation, higher emergency repair costs, and accelerated equipment degradation.

In a case study conducted at a palm oil mill, the boiler station recorded a critically low Overall Equipment Effectiveness (OEE) of 27.03%, far below the world-class benchmark of 85%. Detailed analysis revealed that the primary contributors to this low performance were frequent equipment breakdowns, delayed responses to failures, and the absence of a predictive maintenance approach. Moreover, the reliance on aging equipment without any early-warning mechanisms worsened the situation, resulting in prolonged downtimes and reduced production efficiency.
The adoption of predictive maintenance has proven effective in other industries to address similar challenges. Predictive maintenance utilizes historical and operational data to forecast equipment failures, allowing maintenance teams to intervene before breakdowns occur. This proactive strategy aligns with the principles of Total Productive

Maintenance (TPM) and Industry 4.0, which emphasize data-driven decision-making, digital transformation, and enhanced operational reliability.

Despite the proven benefits of advanced AI techniques, many palm oil mills are constrained by limited data availability and digital infrastructure. Therefore, this study employs logistic regression as the predictive model due to its simplicity, low computational requirements, and transparent outputs that allow maintenance engineers to understand key factors influencing failure risk. Logistic regression offers a practical and cost-effective solution for industrial environments with structured yet limited datasets, such as Excel-based maintenance logs.

This study focuses on the development of an AI-based predictive maintenance system tailored for the boiler station at a palm oil mill. The system enables engineers to upload historical maintenance data, predict failure probabilities for machine components, and receive automated email alerts for components that exceed a predefined risk threshold. By implementing this system, the mill aims to transition from reactive to predictive maintenance, reduce unplanned downtimes, and improve overall equipment reliability and efficiency.

1.1 Objectives

The objective of this study is to assess the Overall Equipment Effectiveness (OEE) of a palm oil mill by examining key indicators such as availability, performance, and quality. Second, it seeks to develop a system predictive maintenance system by utilizing historical maintenance data and implementing logistic regression modeling. Lastly, the study intends to propose a proactive maintenance approach by integrating an alert notification system that can notify engineers in advance of potential failures, enabling timely corrective actions.

2. Literature Review

Total Productive Maintenance (TPM) is a comprehensive maintenance strategy aimed at maximizing equipment effectiveness by involving all levels of the organization in maintenance activities. One of its core metrics is Overall Equipment Effectiveness (OEE), which measures performance through three key indicators: availability, performance, and quality. A world-class OEE value is typically considered to be 85% or higher. However, many studies on palm oil mills in Malaysia report OEE values ranging from 60% to 70%, with frequent breakdowns, long repair durations, and low machine utilization as the primary contributors to inefficiency (Firman et al., 2018; Hairiyah et al., 2019). Despite the promotion of TPM, its implementation in palm oil mills has been inconsistent. Factors such as lack of operator involvement, inadequate maintenance planning, and poor data documentation often hinder the effectiveness of TPM initiatives. Mills that fail to adopt a proactive maintenance culture continue to struggle with low availability and performance rates, particularly in critical stations like boilers.

Traditional maintenance approaches in palm oil mills are predominantly reactive, where machines are repaired only after failures occur. While simple to manage, reactive maintenance results in high unplanned downtime, emergency repair costs, and accelerated equipment degradation. Preventive maintenance strategies offer some improvement by scheduling regular inspections and parts replacements at fixed intervals. However, preventive maintenance does not account for the actual condition of components, leading to unnecessary maintenance or unexpected failures when degradation happens faster than anticipated. These limitations are especially apparent in boiler stations, where components such as burners, furnace grates, and valves are subjected to extreme temperatures and pressures. Without real-time monitoring or predictive insights, maintenance teams often struggle to anticipate failures, resulting in prolonged downtimes and reduced operational efficiency.

Predictive maintenance has emerged as a promising solution that utilizes historical and real-time data to forecast equipment failures before they occur. By predicting when a component is likely to fail, maintenance actions can be scheduled proactively, reducing unplanned downtime and improving resource utilization. The adoption of predictive maintenance aligns with Industry 4.0 principles, which emphasize digital transformation, smart manufacturing, and data-driven decision-making. However, the adoption of predictive maintenance in palm oil mills faces several challenges. Unlike industries with sophisticated digital infrastructures, many palm oil mills still rely on manual data recording methods, such as Excel spreadsheets or handwritten logs. The absence of real-time sensors, IoT-enabled devices, and centralized data systems makes the implementation of advanced AI models impractical. Additionally, the lack of technical expertise and high cost of digitalization further restricts the adoption of complex predictive solutions.

Given the constraints of limited data availability and digital infrastructure in palm oil mills, a practical and interpretable predictive model is required. Logistic regression offers a suitable solution for this context. It is a simple yet effective classification algorithm that can predict the likelihood of component failure based on structured historical maintenance data, such as component type, usage hours, days since last maintenance, and downtime duration. One of the key advantages of logistic regression is its interpretability. Maintenance engineers can easily understand the relationship between input factors and failure probabilities, which builds trust and facilitates data-driven decision-making. Additionally, logistic regression requires fewer data points to train effectively compared to complex models like neural networks, making it ideal for environments where data is limited and collected manually.

Previous studies in the palm oil industry have focused on evaluating machine performance using OEE metrics and recommending general improvements through TPM and preventive maintenance strategies. However, there is a significant gap in translating these analyses into actionable, real-time predictive systems that can proactively guide maintenance activities. Furthermore, existing research has not addressed the practical constraints faced by palm oil mills in adopting advanced AI solutions. This research addresses the gap by developing a predictive maintenance system that leverages logistic regression to predict equipment failures using historical maintenance logs. The system is designed to be practical for mills with limited digital infrastructure, providing failure risk visualizations and automated email alerts to support proactive maintenance planning. By doing so, this study bridges the gap between theoretical maintenance strategies and practical, data-driven solutions tailored for palm oil mills.

3. Methods

This study adopts a data-driven approach to develop a predictive maintenance system specifically for the boiler station in a palm oil mill. The research process began with the collection of historical maintenance data obtained from existing records kept by the mill's maintenance department. The data was structured in Excel spreadsheets and included details such as machine type, component names, failure occurrences, downtime durations, usage hours, and the dates of the last maintenance activities. Before model development, the raw data underwent a thorough cleaning process to remove incomplete, duplicated, and irrelevant entries, ensuring data integrity and reliability.

To enhance the predictive power of the model, feature engineering was performed by creating new variables such as "days since last maintenance," which served as an important indicator of wear and tear. Additionally, categorical variables like component types were encoded into numerical formats suitable for machine learning processing. The target variable was defined as binary, with '1' representing a component failure and '0' indicating normal operation.
Given the structured but limited dataset and the need for an interpretable model, logistic regression was chosen as the predictive algorithm. Logistic regression's simplicity, low computational requirements, and transparent output made it ideal for this industrial environment where data is manually logged and digital resources are limited. The dataset was divided into training and testing subsets in an 80:20 ratio to ensure robust model validation. Model development and training were executed using Python, leveraging the scikit-learn library to optimize model parameters and evaluate performance metrics such as accuracy, precision, recall, and F1-score.

Upon successful model development, the predictive maintenance system was deployed through a web-based application using the Streamlit framework. The interface was designed to be user-friendly, enabling maintenance engineers to upload updated Excel maintenance logs directly into the system. Once uploaded, the system automatically processed the data, ran the failure predictions, and displayed the results in both tabular and graphical formats. Components with high failure probabilities were visualized through bar charts, highlighting critical risk areas that require attention.

An integral part of the system was the implementation of an automated email alert feature. If any component's predicted failure probability exceeded a predefined risk threshold (e.g., 0.90), the system would trigger an immediate email notification to the assigned maintenance engineer. This proactive alert mechanism was designed to ensure timely interventions, preventing potential breakdowns and minimizing downtime.

Finally, the system's predictive accuracy and practical usability were evaluated through testing on real maintenance scenarios. The logistic regression model achieved satisfactory performance, effectively identifying high-risk components while minimizing false alarms. Feedback from maintenance personnel using the system in a live mill environment was also collected to assess the system's user-friendliness and overall impact on maintenance planning.

4. Data Analysis

4.1 Overall Equipment Effectiveness Analysis

After completing the data collection. The next phase of the study involved analyzing the operational effectiveness of the boiler station over a six-month period. This analysis was centered on the calculation of (OEE), a key metric commonly used in industrial settings to evaluate equipment productivity. OEE combines three critical components availability, performance, and quality into a single percentage that reflects the effectiveness of machine operations. Availability refers to the proportion of planned production time during which the machine was actually running. It was calculated by comparing actual operating hours with the total scheduled production hours. Performance measures the speed at which the machine operated relative to its ideal rate, while quality quantifies the percentage of output that met the required standards, excluding rejected or defective units. The OEE for each month was derived using the standard formula in Figure 1.

$$OEE = (Availability \times Performance \times Quality) / 100^2$$

$$Availability = \frac{Actual\ Operationg\ Hours}{Planned\ Production\ Hours} \times 100$$

$$Availability = \frac{400}{720} \times 100\% = 55.56\%$$

$$Ideal\ Output = Actual\ Operating\ Hours\ \times Ideal\ Production\ Rate$$

$$Ideal\ Output = 400\ \times 45 = 18000\ tons$$

$$Performance = \frac{Total\ Output}{Ideal\ Output} \times 100$$

$$Performance = \frac{10000}{18000} \times 100 = 55.56\%$$

$$Quality = \frac{Good\ Output}{Total\ Output} \times 100$$

$$Quality = \frac{9500}{10000} \times 100 = 95\%$$

$$OEE = \frac{Availability\ \times Performance\ \times Quality}{100^2}$$

$$OEE = \frac{55.56 \times 55.56 \times 95}{100^2} = 29.33\%$$

Figure 1. Example calculation of OEE for January.

For instance, in January, the planned production time was 720 hours, but only 400 hours were used for operation, resulting in an availability of 55.56%. The expected ideal output based on a design rate of 45 tons per hour was 18,000 tons, but actual output was only 10,000 tons, giving a performance score of 55.56%. With 500 tons rejected from the 10,000 tons produced, the quality stood at 95.00%. After calculating all of OEE metrics, an OEE is 29.33% for January. This calculation process was repeated from January through June, and the monthly results are summarized in Table 1.

Table 1. Data calculation of OEE metrics and OEE for six months

OEE Metrics				
Month	**Availability (%)**	**Performance (%)**	**Quality (%)**	**OEE (%)**
January	55.56	55.56	95.00	29.32
February	52.08	54.29	95.56	27.02
March	51.02	53.70	94.74	25.96
April	50.00	54.67	94.89	25.94
May	52.42	57.78	95.24	28.85
June	51.39	55.56	94.90	27.09

The OEE values across the six months ranged from 25.94% to 29.32%, significantly lower than the world-class standard of 85%. This trend, illustrated in Figure 2, shows no significant improvement or recovery throughout the period, suggesting persistent performance issues.

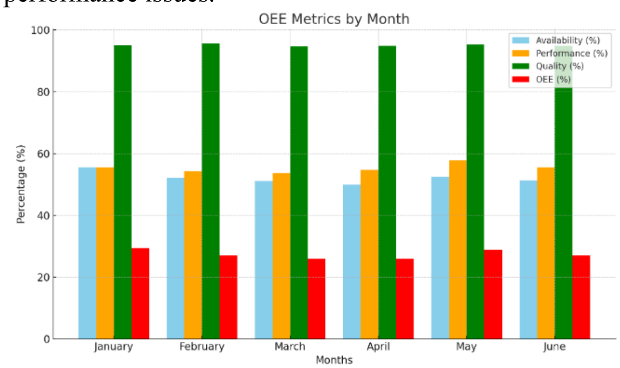

Figure 2. Overall OEE for six months

To further analyze the situation, Figure 3 presents a bar chart that compares the average values of availability, performance, quality and OEE. This chart helps to highlight which part of the machine operation is the weakest and where improvement is most needed. Figure 4.3 the standard of OEE is 85% or shown in red line. Based on the average values shown in the chart, the quality of the product is not a major issue. The average quality score is 95.05%, which means that almost all the products produced meet the required standard. However, both availability and performance show very low values. The average availability is only 52.41%, meaning that the machine was not running for nearly half of the planned production time. This is likely caused by frequent breakdowns, long repair times, or delays in restarting the machine after it stops. The average performance is also low only 55.93%. The performance shows that the machine often runs slower than it should, possibly due to worn parts, unstable operating conditions, or repeated stoppages during production. These two weak factor availability and performance are the main reasons why the overall OEE is so low. These results support the need to improve how maintenance is carried out in the mill.

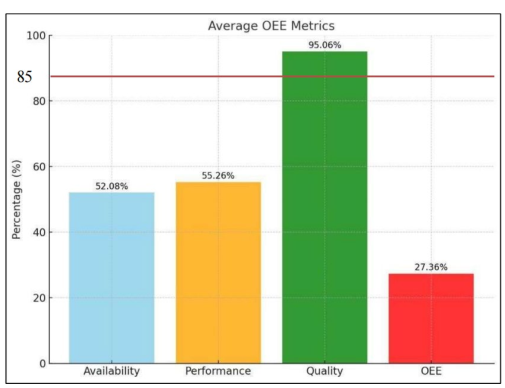

Figure 3. Average of each availability, performance, and quality, and OEE

4.2 Root Cause Analysis

Next, to comprehensively identify the primary factors contributing to the low (OEE) of the boiler station, a root cause analysis (RCA) was performed using cause and effect (Fishbone) diagrams. The analysis specifically targeted the two OEE components, availability and performance which demonstrated the most significant losses in the operation of the boiler station. Although the quality component maintained a consistently high level, inefficiencies in availability and performance were identified as critical bottlenecks that required immediate attention. By systematically examining the factors influencing these two OEE pillars, the RCA aimed to isolate and prioritize the underlying mechanical, operational, and maintenance-related issues that hindered the boiler station's overall productivity.

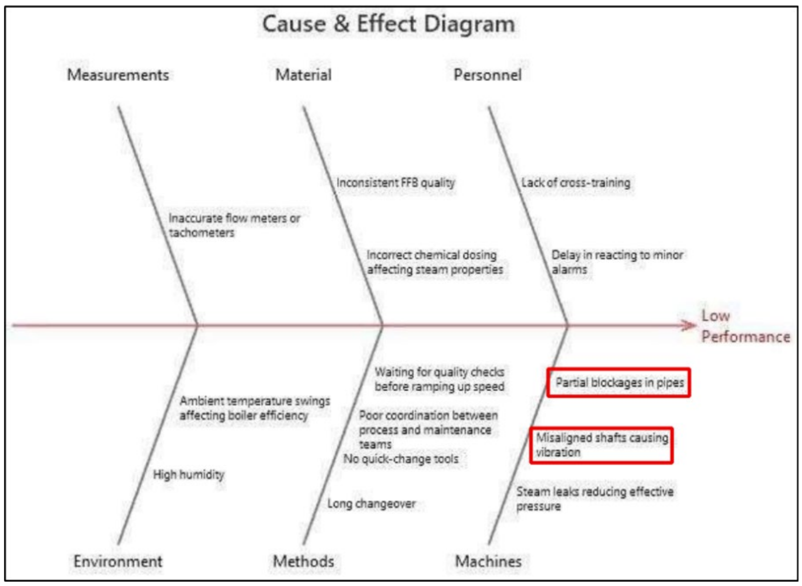

Figure 4. Cause and effect diagram (low performance)

The Cause-and-Effect Diagram developed to analyze low performance as shown in Figure 4, revealed multiple interrelated factors contributing to the boiler's inability to operate at its designed capacity. Among these, two root causes were identified as having the most direct and severe impact on the performance metric. The first is partial blockages in pipes, a recurring issue stemming from the accumulation of deposits, sludge, and debris within the pipeline system. These blockages restrict the flow of steam and water, leading to unstable pressure levels and irregular heat transfer efficiency. As a result, the boiler cannot sustain its intended output rate, which forces operators to reduce operational loads to prevent system overload or damage. This phenomenon leads to a consistent reduction in the actual performance rate compared to the theoretical maximum, severely impacting the performance ratio in the OEE calculation. The second critical cause is misaligned shafts causing excessive vibration within the rotating machinery. Shaft misalignments, whether due to improper installation, wear, or lack of precision in maintenance, lead to abnormal vibrations that compromise the mechanical stability of the equipment. These vibrations not only reduce energy efficiency but also necessitate frequent manual interventions, such as speed adjustments and equipment inspections, to mitigate potential damage. Over time, this misalignment problem limits the boiler's ability to operate at optimal speed and consistency, thereby reducing its performance output relative to design specifications. Both of these issues, pipe blockages and mechanical misalignments force the boiler to operate below its maximum efficiency threshold, directly impairing the performance component of OEE. Addressing these factors is essential to restoring stable and high-performance operations.

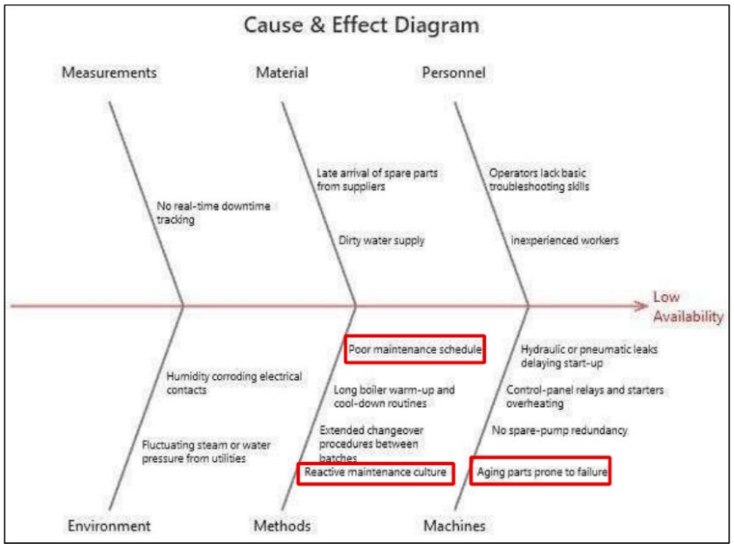

Figure 5. Cause and effect diagram (low availability)

In parallel, a root cause analysis was conducted focusing on low availability as shown in Figure 5, where several factors were found to contribute to excessive downtime and reduced equipment uptime. The most prominent among these were related to systemic gaps in maintenance practices and equipment lifecycle management. A key contributor identified is the absence of a structured maintenance schedule. Currently, maintenance activities are largely performed in a reactive manner, repairs and inspections are initiated only after failures or malfunctions occur. This lack of proactive scheduling has resulted in prolonged unplanned downtimes, where repairs take longer due to a lack of preparation, spare parts, or personnel allocation. The absence of a preventive framework means that minor issues, which could have been addressed during planned downtimes, are left to escalate into critical failures. Closely tied to this is the reactive maintenance culture that dominates current operational practices. This approach, which emphasizes fixing problems only after they arise, has made the boiler system highly susceptible to unexpected breakdowns. Each instance of reactive repair disrupts production schedules and introduces additional delays, as failures are often detected only after they have caused significant equipment damage. This not only prolongs repair times but also increases the frequency of emergency downtimes, further reducing the system's availability. Additionally, aging parts prone to failure were recognized as a critical mechanical factor contributing to low availability. Many components within the boiler system, including pumps, valves, and mechanical seals, have surpassed their designed operational lifespan. Continued usage of these deteriorating parts increases the likelihood of sudden breakdowns. The absence of a systematic parts replacement policy exacerbates this issue, leading to higher machine unavailability as frequent failures require extended repair or replacement efforts. Collectively these root causes, poor maintenance scheduling, a reactive repair approach, and aging components are directly responsible for the reduction in operational uptime, significantly lowering the Availability metric of the boiler's OEE. Addressing these systemic weaknesses is crucial to restoring continuous and reliable boiler operations.

4.3 Solution Proposal: System Predictive Maintenance

Following the Root Cause Analysis, it was evident that the primary factors contributing to the low Overall Equipment Effectiveness (OEE) of the boiler station were related to mechanical failures, lack of proactive maintenance scheduling, and reliance on reactive repairs after breakdowns occurred. Traditional maintenance methods, including scheduled inspections, were insufficient to predict imminent failures, resulting in frequent unplanned downtimes and reduced machine performance. To address these issues, this study proposes the development of a predictive maintenance system. The objective of this system is to predict potential failures using historical maintenance and operational data, enabling engineers to take preventive actions before actual breakdowns occur. By shifting from reactive to predictive maintenance, the system aims to minimize downtime, enhance component reliability, and ultimately improve the boiler station's availability and performance metrics. The solution leverages logistic regression as its core predictive algorithm, providing a data-driven method to assess the probability of component failures based on past failure patterns, usage hours, and maintenance history. In addition, the system integrates a user-friendly dashboard and automated email notification feature to ensure that maintenance teams receive timely alerts when components are at high risk. This predictive approach aligns with Total Productive Maintenance (TPM) principles and supports the palm oil mill's journey towards digital transformation and Industry 4.0 readiness.

4.3.1 System Development

The development of the predictive maintenance system followed a structured approach to meet the operational needs of the palm oil mill's boiler station. The system aimed to predict probabilities of component failure using historical maintenance records, enabling proactive preventive actions. Python was used as the development platform due to its flexibility and strong machine learning libraries, with Pandas and NumPy for data preprocessing and Scikit-learn for training the logistic regression model. Logistic regression was chosen for its simplicity, interpretability, and effectiveness in binary classification tasks. A system flowchart was designed to guide the workflow, starting from data upload, preprocessing, prediction, classification, and triggering email alerts if a component's failure probability exceeded 0.90. This flowchart (Figure 6) structured the development and explanation of each system function in subsequent sections.

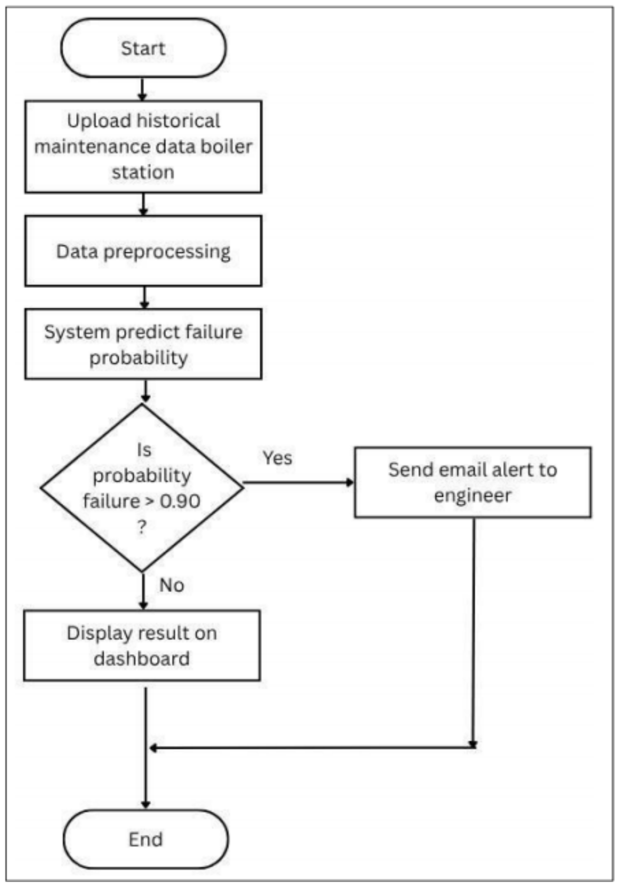

Figure 6. Flowchart of the system predictive maintenance

4.3.1.1 Upload historical maintenance data and preprocessing

The first step in the predictive maintenance system is uploading historical maintenance data, which serves as the core input for preprocessing and failure risk assessment. The system accepts Excel (xlsx) files, ensuring compatibility with existing maintenance workflows. Key fields required in the dataset include Component Name, Last Maintenance Date, Usage Hours, Downtime, and Failure Status (1 for failure, 0 for no failure). These values are derived from actual maintenance logs manually recorded by technicians during inspections as shown in Table 2. The system dashboardas shown in Figure 7, built using Streamlit, provides a user-friendly interface for uploading the data. Upon upload, the system automatically validates the file structure, ensuring required columns are present. If errors are detected, users are prompted to correct and re-upload the file. This process ensures accurate data entry while being accessible to maintenance personnel with no programming knowledge. The data preprocessing stage ensures that uploaded maintenance data is clean, consistent, and ready for use by the logistic regression prediction model. Raw data often contains incomplete fields, text values, or inconsistent formats, which must be standardized. The system converts the "Last Maintenance Date" into a numerical "Days Since Last Maintenance" feature, making it usable for prediction. Categorical data, such as Component Names, are encoded into numerical labels through label encoding. The system also checks for missing or inconsistent entries, removing incomplete rows or filling gaps with average values as needed. Key features like Component Type, Usage Hours, Downtime Duration, and Days Since Last Maintenance are selected for prediction, as they are critical indicators of component health. This entire preprocessing workflow is automated, requiring no manual intervention from users. Once data is cleaned and transformed, it proceeds directly into the logistic regression model for failure probability calculation, ensuring the model receives high-quality, structured data to improve prediction accuracy.

Table 2. historical maintenance data

Date	Component	Last Maintenance	Failure	Days Since Last Maintenance	Usage Hours	Downtime (min)
2025-02-10	Economizer Tubes	2024-12-14	Yes	58	1744	96
2025-03-18	Air Preheater	2025-02-16	No	30	2093	0
2025-01-29	Economizer Tubes	2024-12-13	No	47	1311	0
2025-02-17	Feedwater Pumps	2025-01-14	No	34	1819	0
2025-01-09	Burner	2024-11-23	Yes	47	1895	149
2025-03-21	Feedwater Pumps	2025-02-28	No	21	1253	0
2025-02-24	Furnace Grate	2025-01-29	Yes	26	1781	141
2025-01-16	Safety Valve	2024-12-23	No	24	1376	0
2025-02-12	Feedwater Pumps	2024-12-22	No	52	2397	0

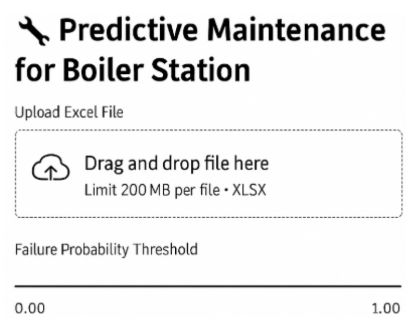

Figure 7. Dashboard of the system

4.3.1.2 Model selection and justification

Logistic regression was chosen as the predictive model for this system due to its simplicity, interpretability, and suitability for binary classification tasks, specifically predicting whether a component will fail (1) or not fail (0). Unlike complex AI models, logistic regression is effective for structured industrial datasets with limited records, which is common in environments lacking real-time sensor data. Its fast training, transparent outputs, and minimal computational requirements make it ideal for integration into a web-based interface. Logistic regression provides clear probability scores that maintenance personnel can easily interpret to identify high-risk components for preventive action. This model strikes a balance between predictive performance, ease of use, and practical application, ensuring the system remains accessible to non-technical users while supporting proactive maintenance decision-making.

4.3.1.3 Logistic regression analysis

Following the model selection and justification, the next step in the system development process was the creation and training of the logistic regression model. Logistic regression is a supervised learning algorithm used to predict the probability of a binary outcome in this case, whether a machine component is likely to fail (1) or not (0). The model was trained using preprocessed historical maintenance data containing labeled outcomes based on actual past failures.

The training process began with selecting the appropriate input features from the dataset. These included the encoded component type**,** usage hours**,** downtime (minutes), and days since last maintenance. Each of these features contributes important information related to the wear, usage, and operational stress of the equipment. The target variable, or output, is the failure status, which indicates whether the component has experienced a failure (1) or not (0) in the past. The logistic regression model operates by estimating the probability of failure using a linear combination of input features, passed through a sigmoid function. The sigmoid function maps the linear combination to a value between 0 and 1, which represents the likelihood that the component will fail. The mathematical representation of the model is shown in the equation below:

$$P(Y=1|X) = \frac{1}{1+e^{-z}}$$

$$where\ z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

In this equation, P(Y=1|X) represents the probability that a failure will occur based on input features X1, X2, …, Xn, and β0 to βn are the model coefficients. After training, the model generates a probability score between 0 and 1 for each component. If the score exceeds a defined threshold set at 0.90 in this project the component is flagged as high-risk, and preventive actions are recommended. To illustrate how the prediction process works, consider an actual data entry from the collected maintenance records on Table 3. On 9 January 2025, a burner component was recorded with 47 days since its last maintenance, 1895 total usage hours, and 149 minutes of downtime. The component type "Burner" was numerically 68 encoded as 1 for use in the model. Assume that the logistic regression model produced the following coefficients after training:

- Intercept (β_0) = -2.5
- β_1 (Days Since Last Maintenance) = 0.04
- β_2 (Usage Hours) = 0.002
- β_3 (Downtime in minutes) = 0.01
- β_4 (Component Type) = 0.3

By substituting these values into the logistic regression equation, the linear combination is calculated as follows:

$$z = -2.5 + (0.04 * 47) + (0.002 * 1895) + (0.01 * 149) + (0.3 * 1)$$
$$z = -2.5 + 1.88 + 3.79 + 1.49 + 0.3 = 4.96$$

Next, the sigmoid function is applied to convert this result into a probability:

$$P(Y=1) = \frac{1}{1+e^{-4.96}} = 0.993$$

This yields a failure probability of approximately 99.3%. Since this value exceeds the defined threshold of 90%, the system classifies this component as high risk. As a result, it is highlighted in the dashboard and an email notification is automatically sent to the maintenance engineer recommending preventive action. This example clearly demonstrates how the trained logistic regression model utilizes historical maintenance data to make accurate predictions about future failures. The model's interpretability and low computational cost make it suitable for real-world applications in palm oil mill operations, especially in resource-constrained environments. By integrating this model into the maintenance system, engineers are equipped with timely and data-driven insights to reduce downtime and improve operational efficiency

5. Results and Discussion

5.1 Result Analysis

One of the core outputs of the predictive maintenance system is the prediction table, which is generated immediately after the maintenance data is uploaded and processed. This table provides a detailed summary of each boiler component's operational condition and associated failure risk, allowing maintenance teams to make quick and informed decisions. As shown in Table 3, this was the output from the system after upload the historical maintenance data.

Table 3. Prediction table output

	Date	Component	Last Maintenance	Failure	Days Since Last Maintenance	Usage Hours	Downtime (min)	Probability failure
0	10/2/2025	Economizer Tubes	14/12/2024	1	58	1744	96	0.62
1	18/3/2025	Air Preheater	16/2/2025	0	30	2093	0	0.14
2	29/1/2025	Economizer Tubes	13/12/2024	0	47	1311	0	0.14
3	17/2/2025	Feedwater Pumps	14/1/2025	0	34	1819	0	0.36
4	9/1/2025	Burner	23/11/2024	1	47	1895	149	0.91
5	21/3/2025	Feedwater Pumps	28/2/2025	0	21	1253	0	0.27
6	24/2/2025	Furnace Grate	29/1/2025	1	26	1781	141	0.91
7	16/1/2025	Safety Valve	23/12/2024	0	24	1376	0	0.1
8	12/2/2025	Feedwater Pumps	22/12/2024	0	52	2397	0	0.02

Following the insights provided by the prediction table, the system further enhances its practicality through an automated email alert notification feature, ensuring that critical predictive insights are communicated in real-time to maintenance personnel. While the dashboard and prediction tables allow engineers to visually monitor component conditions within the system, the email alert mechanism ensures that high-risk warnings are actively pushed to the relevant personnel, even when they are not logged into the system. Once the system completes the prediction and classification process, it automatically scans for components with a failure probability exceeding the predefined threshold of 0.90. If such components are detected, the system triggers an automated email alert directed to the assigned maintenance engineer. This alert contains essential information, including the component name, the predicted failure probability, and a clear recommendation indicating that immediate inspection or maintenance is required.
The inclusion of both the component's identity and its associated risk score ensures that engineers can quickly assess the urgency of the situation. The component name focuses attention on the exact unit needing intervention, while the failure probability provides a quantitative measure of risk severity. This allows engineers to make informed, data-driven decisions without the need to immediately access the system dashboard. An example of the automated email notification format is illustrated in Figure 8, which shows how critical information is structured to facilitate fast and effective maintenance response.

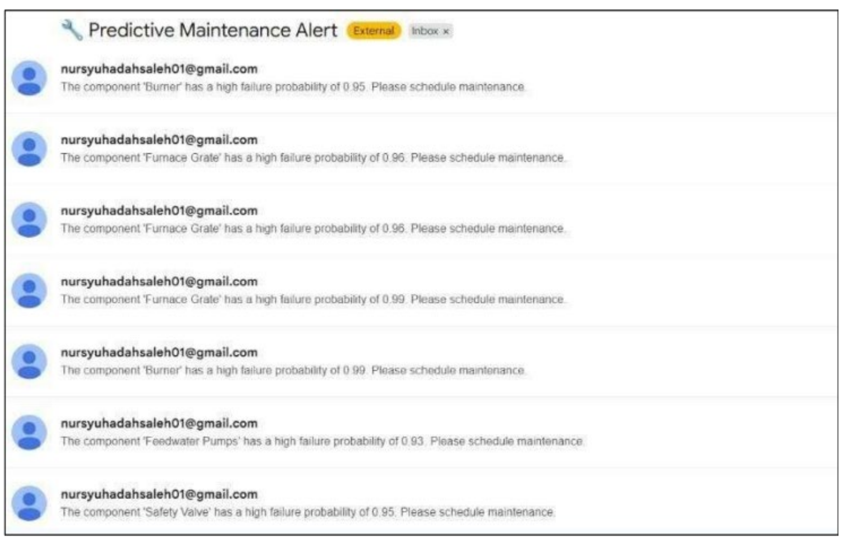

Figure 8. Automated Email Alert for High-Risk Component

The system uses SMTP (Simple Mail Transfer Protocol) to send emails. Once a component's failure probability exceeds the set threshold (0.90), the system automatically sends an email with the component name, failure probability, and a message recommending urgent maintenance. The backend is programmed using Python's smtplib and email libraries, which handle the connection to an external mail server (e.g., Gmail) and send the alert within seconds. This email function works automatically; no extra steps are needed from the user. When an engineer uploads a maintenance file and a high-risk component is found, the alert is sent immediately. This helps ensure that maintenance teams are aware of critical risks without relying on manual checks, allowing for quicker response and better coordination, whether the engineer is at the plant or off-site.

By sending alerts in real-time, the system helps engineers take preventive actions faster, reduces the risk of unexpected failures, and supports the goal of achieving predictive maintenance in daily operations.

5.2 Result Validation

After the development and deployment of the predictive maintenance system, a validation process was conducted to ensure that the system performs reliably in both prediction accuracy and operational functionality. This step was crucial to confirm that the model's predictions are meaningful and that the system can be used effectively in an actual industrial environment. From a predictive performance standpoint, the logistic regression model was trained and tested using historical maintenance data labelled with actual failure outcomes. The dataset was divided into training and testing subsets, and the model's performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The results of the classification report for the test dataset are presented in Table 4.

Table 4. Classification report of the logistic regression model on test data

Class	Precision	Recall	F1-Score	Support
0 (No failure)	0.67	0.92	0.77	13
1 (Failure)	0.92	0.65	0.76	17
Accuracy	-	-	0.77	30
Macro Average	0.79	0.79	0.77	30
Weighted Average	0.81	0.77	0.77	30

The predictive maintenance model achieved an overall accuracy of 77%, with a high precision of 0.92 for predicting failures, ensuring reliable alerts with few false positives. However, the recall was 0.65, meaning some actual failures were not detected, which is acceptable for this initial system prioritizing precision in high-risk scenarios. Functional testing confirmed that all system modules file upload, prediction logic, risk classification, and email alerts operated correctly. Feedback from engineering supervisors validated that the dashboard outputs were clear and helpful for maintenance planning. Overall, the system was proven to perform reliably, providing accurate predictions and actionable insights, making it suitable for real-world implementation in palm oil mill operations to improve maintenance planning and early failure detection.

5.3 Discussion

The predictive maintenance system developed in this study addresses the need for intelligent, data-driven maintenance in palm oil mills, aiming to reduce unplanned downtime and improve Overall Equipment Effectiveness (OEE). Using Python and Streamlit, the system applies a logistic regression model to predict component failures based on historical maintenance data. Logistic regression was chosen for its simplicity, transparency, and suitability for environments where large real-time datasets are unavailable. The system achieved 77% accuracy and 92% precision, ensuring reliable high-risk predictions, although some failures were missed due to a recall of 65%. A key feature is the adjustable probability threshold (default 0.90), giving engineers flexibility to balance between sensitivity and false positives. Outputs such as the prediction table and automated email alerts help engineers prioritize maintenance tasks and respond quickly to critical risks without constant dashboard monitoring. Validation confirmed the system's reliability in processing data, predicting failures, and sending real-time alerts. However, limitations include dependency on historical data quality, manual data uploads, and the logistic regression model's inability to capture complex patterns. Future improvements should focus on automating data collection and exploring advanced predictive models. In conclusion, the system successfully shifts maintenance practices from reactive to predictive, offering a practical and interpretable solution for palm oil mill operations, and represents an important step towards digital transformation in maintenance management.

5.4 Recommendations for Industry Application

To maximize the impact of the developed predictive maintenance system, several improvements are recommended for future industrial application. First, the system should be expanded beyond the boiler station to include other critical areas like sterilizer, press, clarification, and kernel stations. This would enable mill-wide monitoring and more effective maintenance coordination. Second, integrating real-time sensor data such as temperature, vibration, and pressure can enhance prediction accuracy and enable earlier failure detection.

Besides, the system should also support automatic model retraining to adapt to new maintenance data and ensure continuous learning. Additionally, implementing multi-channel notifications (e.g., WhatsApp or SMS) can improve response time by reaching engineers faster than email. Lastly, enhancing security through user logins and role-based access control will protect data integrity and improve usability for different personnel levels. These recommendations aim to support a smarter, scalable, and Industry 4.0-ready maintenance system in palm oil mills.

6. Conclusion

This study successfully developed a predictive maintenance system to address low equipment effectiveness at a palm oil mill's boiler station by shifting from reactive to proactive maintenance strategies. By analyzing six months of operational data, the system identified low availability and performance as key factors reducing Overall Equipment Effectiveness (OEE) to 27.03%. A logistic regression model was built using historical maintenance records to predict failure probabilities of machine components, with high-risk items flagged for preventive action. The system, deployed via a user-friendly web interface, achieved 77% accuracy and 92% precision in predicting failures, ensuring reliable identification of critical risks. Additionally, an automated email notification feature was integrated to alert engineers in real-time when components exceed risk thresholds, enabling quicker interventions. Overall, the system provides a practical, interpretable solution that empowers engineers with data-driven insights, enhances maintenance planning, reduces unplanned downtime, and contributes directly to improving OEE, aligning with Industry 4.0 and TPM initiatives.

References

Analisis efektivitas mesin expeller dengan implementasi TPM (Total Productive Maintenance) di PT. Wilmar Nabati Indonesia Gresik. *Kaizen: Management Systems & Industrial Engineering Journal*, 2(2), 76–77, 2019.

Diniaty, D., Susanto, R., & Jurusan Teknik Industri Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau. Analisis Total Productive Maintenance (TPM) pada stasiun kernel dengan menggunakan metode Overall Equipment Effectiveness (OEE) di PT. Surya Agrolika Reksa. *Jurnal Teknik Industri*, 3(2), 2017.

Firman, 1, Thabrani, G., 1, Violeta, V. P., 1, & Manajemen, Universitas Negeri Padang, Padang, Indonesia. Analisis peningkatan kinerja pemeliharaan mesin dengan Total Productive Maintenance (TPM) pada mesin boiler pabrik kelapa sawit PT. Perkebunan Nusantara VI unit usaha Rimbo Dua Tebo-Jambi. *Jurnal Kajian Manajemen Bisnis*, 8(2), 55–65, 2018. http://ejournal.unp.ac.id/index.php/jkmb.

Jayatillake, K., & Amirthalingam, K. Overall equipment effectiveness in Sri Lankan palm oil industry: A case study of ABZ Company. Master's thesis, GSJ. *Faculty of Graduate Studies & University of Colombo, Master in Manufacturing Management 2020/2021*, 10(4), 69, 2021. https://www.globalscientificjournal.com.

Latif, A., Purnomo, R., & Program Studi Teknik Industri Sekolah Tinggi Teknologi Industri (STTIND) Padang. Analisis Total Productive Maintenance (TPM) menggunakan Overall Equipment Effectiveness di PT. Perkebunan Nusantara VI Ophir. *Jurnal Sains dan Teknologi*, 19(2), 86, 2019.

Measuring overall equipment effectiveness (OEE) palm oil mill in Indonesia. *International Journal of Recent Trends in Engineering and Research*, 3(12), 164–170, 2017. https://doi.org/10.23883/ijrter.2017.3551.zq21l.

Measuring OEE in Malaysian palm oil mills. *Interdisciplinary Journal of Contemporary Research in Business*, 4(2), 733–735, 2012. https://ijcrb.webs.com.

Supriyadi, S., Ramayanti, G., & Afriansyah, R. Analisis Total Productive Maintenance dengan metode Overall Equipment Effectiveness dan Fuzzy Failure Mode and Effects Analysis. *Sinergi*, 21(3), 165, 2017. https://doi.org/10.22441/sinergi.2017.3.002.

Susilawati, A., Tasri, A., & Arief, D. A framework to improve equipment effectiveness of manufacturing process: A case study of pressing station of crude palm oil production, Indonesia. *IOP Conference Series: Materials Science and Engineering*, 602(1), 012041, 2019. https://doi.org/10.1088/1757-899x/602/1/012041.

Wan Mahmood, W. H., Abdullah, I., Md Fauadi, M. H. F., Ab Rahman, M. N., Sustainable and Responsive Manufacturing Research Group, & Advanced Manufacturing Research Group. OEE measures for sustainable environment in palm oil mill: A review. *Sci. Int. (Lahore)*, 26(5), 1855–1859, 2014.

Biographies

Nursyuhadah Binti Saleh is a Bachelor of Mechanical (Industrial) Engineering student at Universiti Teknologi Malaysia, Johor Bahru Malaysia. Have experience as a maintenance engineer intern at Brantian Palm Oil Mill Tawau, Sabah.

Jafri Mohd Rohani is an Associate Professor at the Faculty of Mechanical Engineering, Universiti Teknologi Malaysia (UTM), with over 30 years of experience teaching industrial engineering courses. He holds a PhD from UTM (2014), an MSc in Industrial Systems Engineering from Ohio University (1995), and a BSc in Industrial Engineering from New Mexico State University (1988). He has supervised numerous PhD and Master's students and is active in research, with over 22 Web of Science and 10 Scopus-indexed journal articles, a research book, and two book chapters. He has secured more than RM 1 million in research funding and consultancy projects. His scholarly work has earned him an h-index of 14 on Scopus and 19 on Google Scholar, with over 1,747 citations. Dr. Rohani has been invited as a keynote speaker at international conferences and serves on editorial boards and technical committees for various journals and conferences. Additionally, he consults for NIOSH Malaysia and DOSH on occupational safety and health projects and coordinates several MOUs with institutions such as NIOSH Malaysia and Universitas Sebelas Maret, Surakarta.

Nurul Rawaida Ain Burhani is a Senior Lecturer at the Faculty of Mechanical Engineering, Universiti Teknologi Malaysia (UTM) specializing in reliability and the engineering application of Artificial Intelligence. She holds a PhD in Mechanical Engineering from Universiti Teknologi Petronas and has extensive teaching experience in various engineering disciplines. She has been recognized with multiple awards, including the QIU Innovative Teaching and Learning Award and the Best Paper Presentation at flagship conference. She has secured significant research grants and has contributed to various research publications on predictive maintenance. In addition to her academic role, she actively participates in NGOs: YOSH MyFundAction and National STEM Association, youth development programs, and STEM initiatives. She has also been a reviewer for prestigious journals and conferences and has been invited to share her expertise as a speaker and trainer in various academic and professional settings worldwide.

Puteri Mawardati Binti Yasir is a Master of Science (Industrial Engineering) student at Universiti Teknologi Malaysia, Johor, Malaysia. Have three years of experience as an Industrial Engineer during her time working in a contract manufacturing company in Seberang Jaya, Penang. Decided to enroll into Master degree to enhance knowledge and the technical understanding of industrial engineering.