Jurnal Dinamika Vokasional Teknik Mesin
Vol. 10, No. 2, October 2025, pp. 120-134
https://journal.uny.ac.id/index.php/dynamika/issue/view/89737
ISSN: 2548-7590, DOI: 10.21831/dinamika.v10i2.89737

Prescriptive Maintenance of Komatsu Dump Truck HD785-7 Using Naïve Bayes Classifier Under Full Maintenance Contract

Paimin*, Gembong Baskoro, Iwan Sahputra
Master of Mechanical Engineering, Swiss German University, Tangerang City, Indonesia.

Article Info

Article history:

Received Sep 15,2025
Revised Sep 16, 2025
Accepted Sep 18, 2025
Published Oct 30, 2025

Keywords:

Prescriptive Maintenane (RxM)
Physical Availability (PA)
Full Maintenance Contract (FMC)
Naïve Bayes Classifier
Data Mining
Predictive Analytics

ABSTRACT

This study proposes a Prescriptive Maintenance (RxM) framework aimed at improving the Physical Availability (PA) of Komatsu Dump Trucks HD785-7 operated under a Full Maintenance Contract (FMC) at PT ABC Site Sangkulirang. The research integrates the DMAIC methodology with the Knowledge Discovery in Databases (KDD) process to systematically analyze operational failures. Historical breakdown data were preprocessed and modeled using a Naïve Bayes (NB) classifier, selected for its robustness in handling categorical features common in maintenance records. The model demonstrated high predictive performance with 97.93% accuracy, 100% precision, 94.12% recall, and an AUC of 0.995, indicating strong reliability in distinguishing high-risk conditions. The RxM framework was embedded into daily maintenance planning and Standard Operating Procedures (SOPs), supported by a monitoring dashboard for continuous feedback and retraining. As a result, the proportion of Breakdown Unscheduled (BUS) events decreased from 45% in 2024 to 26% in mid-2025, while fleet PA consistently exceeded the contractual target of 92%, reaching 95.5%. These findings confirm that embedding prescriptive analytics into maintenance workflows not only reduces unplanned downtime but also enhances resource allocation and decision-making. The case study highlights the practical value of combining statistical learning with structured process improvement to drive digital transformation in mining operations.

Corresponding Author:

Paimin
Master of Mechanical Engineering, Faculty of Engineering and Information Technology
Swiss German University
Tangerang, Banten 15143 - Indonesia
Email: paimin.paimin@student.sgu.ac.id

INTRODUCTION

The reliability of heavy equipment plays a crucial role in the mining sector, as downtime directly affects productivity and operational costs. PT ABC operates a fleet of Komatsu Dump Truck HD785-7 under an FMC, which requires a minimum PA of 92%. However, in 2024, the average PA of the fleet reached only 90.6%, with a high rate of BUS at 45%.

With advancements in technology, maintenance strategies have evolved significantly. Initially, the approach was Reactive Maintenance (RM), where equipment was only repaired after failure. Over time, this evolved into Preventive Maintenance (PM), focusing on scheduled maintenance to prevent failures through component replacements or routine maintenance. Subsequently, Condition-Based Maintenance (CBM) emerged, relying on equipment condition monitoring to detect potential issues based on specific metrics. Predictive Maintenance (PdM) followed, using sensor data and analytics to predict failures before they occur, though it was still limited in providing specific action recommendations. Finally, RxM integrated more complex analytics and diverse data to not only predict failures but also offer actionable recommendations to prevent these failures, such as scheduling part replacements or operational adjustments. This approach allows for more targeted decision-making and enables maintenance teams to act faster and more efficiently in maintaining optimal equipment performance.

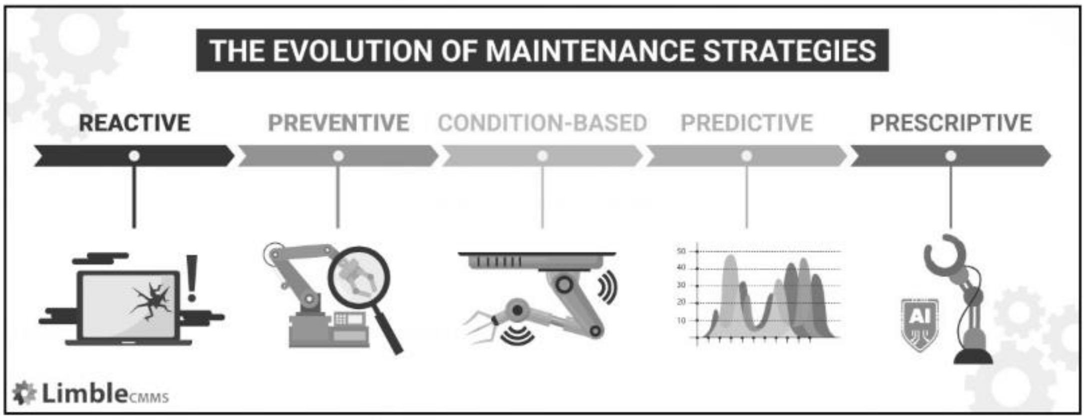

Figure 1. Evolution of Maintenance Strategies (Source: Limble, 2025)

The evolution of maintenance strategies, as shown in Figure 1, demonstrates the shift from reactive maintenance to data-driven and more sophisticated analytical strategies. PdM uses analytics to predict potential failures, RxM goes further by providing more detailed and specific steps to address detected issues. By integrating larger datasets and advanced analysis techniques, RxM facilitates more proactive and efficient maintenance decisions, ensuring quicker actions to prevent severe equipment failures (Limble, 2025; Pillai et al., 2021).

Several previous studies have laid the foundation for the approach used in this research. Ferdinand (2021) emphasized the importance of prioritizing emergency maintenance needs to prevent unscheduled failures in heavy equipment operations managed under FMC. Although a considerable body of research has focused on the application of PdM to anticipate equipment failures, the comprehensive implementation of RxM frameworks in real-world industrial environments remains relatively scarce (Choubey et al., 2021). To bridge this gap, recent investigations have advanced the methodological approaches in heavy mining equipment maintenance. For instance, Moniri-Morad and Sattarvand (2023) compared various system reliability evaluation methods in mining dump trucks, highlighting how reliability modeling contributes to identifying critical failure modes and supporting optimized maintenance planning. Complementing this perspective, Rahimdel et al. (2024) employed a dynamic Bayesian network approach to analyze the reliability of mining truck subsystems, providing empirical evidence that probabilistic models can enhance decision-making in maintenance scheduling and align with the prescriptive paradigm. Similarly, Giacotto, Marques, and Martinetti (2023) proposed a holistic and scalable optimization framework for prescriptive maintenance that leverages IoT, asset health prognostics, and advanced analytics, demonstrating the applicability of RxM in complex and dynamic industrial settings.

One of the techniques used in RxM is the NB algorithm. This algorithm has proven effective in classification tasks, especially for categorical data commonly encountered in industrial equipment maintenance. The strength of NB lies in its ability to handle relatively small datasets with probabilistic results that are easy to interpret, making it highly useful for maintenance decision-making (Choubey et al., 2020). Several previous studies have applied NB in predictive contexts to forecast equipment failures, reporting high classification accuracy rates (Chazhoor et al., 2021; Fajar et al., 2021). This study adapts NB into the RxM cycle, focusing on maintenance operations under the FMC and linking failure predictions with appropriate maintenance actions.

Moreover, the concept of RxM has advanced with the integration of sophisticated tools. For example, Mao et al. (2023) combined RxM with digital twin technology to optimize maintenance scheduling under resource constraints. This study builds on these insights but focuses on practical application in heavy mining equipment, using NB to drive the integrated RxM program within maintenance operations.

In terms of algorithms, NB classification has proven effective for maintenance data analysis. Chazhoor et al. (2020) compared NB with other techniques for predicting machine failures and found that NB strikes a strong balance between simplicity and accuracy, particularly for categorical data and imbalanced datasets. Fajar et al. (2021) applied NB to production machine maintenance logs and reported that NB's ability to handle categorical features and incomplete data resulted in reliable predictions in the industrial context. These findings support the choice of NB for classifying maintenance needs in this study.

RapidMiner is an open-source data science platform that facilitates the entire data analysis process, from preparation to deployment. With its user-friendly graphical interface, RapidMiner enables non-technical users to build and execute machine learning workflows, including Naïve Bayes for classification. The RapidMiner Altair AI Studio application provides an intuitive workspace for designing and managing analytical processes, making it an effective tool for advanced data-driven solutions.

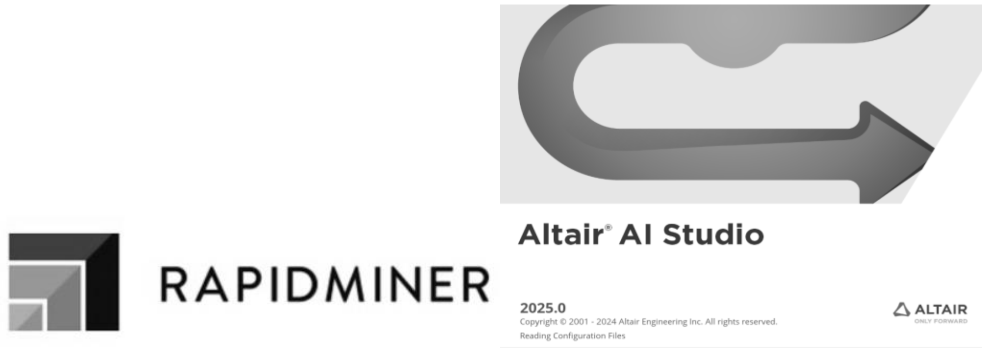

Figure 2. RapidMiner Altair AI Studio

To ensure that the implementation of RxM delivers sustainable impact, it is essential to integrate RxM with a structured improvement framework. One widely used framework is DMAIC, which provides a systematic approach to analyzing and improving processes. The literature emphasizes the importance of combining data-driven models with structured improvement frameworks to ensure that the changes made are not only temporary but can result in long-term improvements (Lumban Raja et al., 2024).

This study aims to apply RxM, supported by NB classification, to identify patterns in historical failure data and provide preventative action recommendations. The uniqueness of this research lies in the integration of RxM into FMC operations through updated SOPs and the creation of a dashboard, which aims to close the gap between prediction and data-driven decision-making.

METHOD

Research Framework

This research follows the DMAIC (Define-Measure-Analyze-Improve-Control) methodology integrated with the KDD (Knowledge Discovery in Databases) workflow. The DMAIC phases are used to define the problem, collect data, analyze results, implement improvements, and control outcomes to ensure continuous improvement. Meanwhile, the KDD process supports this research by extracting insights from the existing data through selection, cleaning, and modeling relevant data. The flow diagrams illustrating the research framework are shown in Figures 3 and 4.

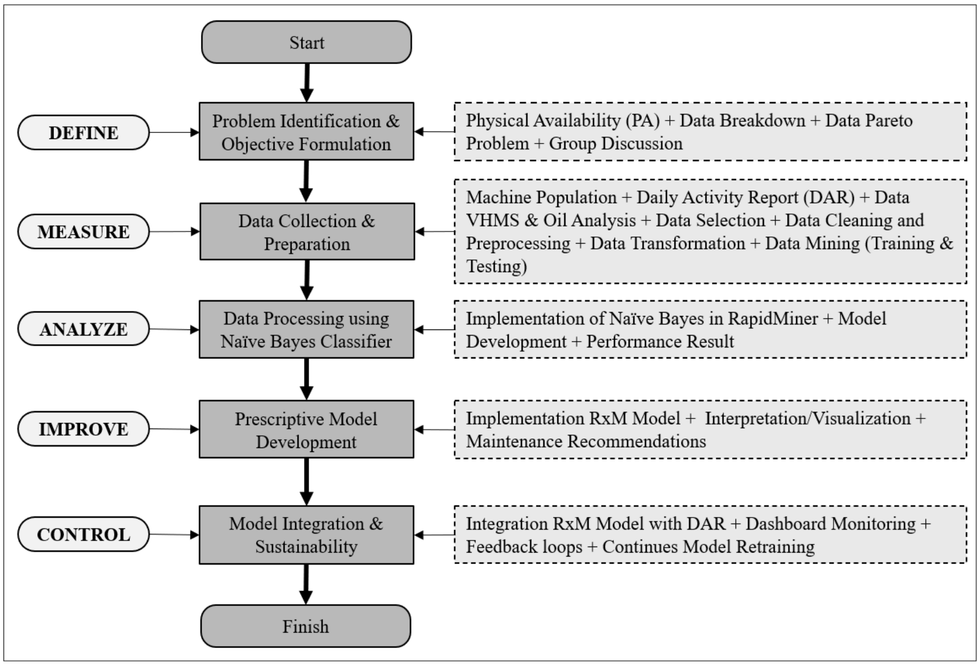

Figure 3. DMAIC Framework

This diagram illustrates the DMAIC phases used in this research. The first phase, Define, aims to define the existing problem, such as low PA of the fleet. Measure involves gathering relevant operational failure data. Next, in the Analyze phase, the collected data is analyzed to identify failure patterns. During the Improve phase, prescriptive maintenance solutions are applied to enhance PA, and the final Control phase ensures that the improvements are maintained, and results are monitored over time.

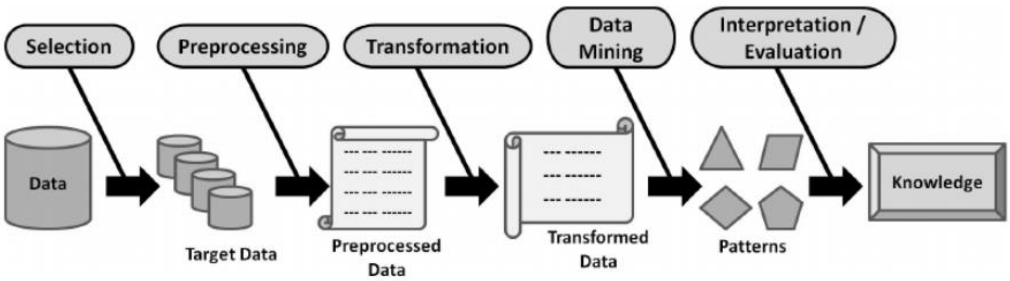

Figure 4. Knowledge Discovery in Databases workflow

This diagram shows the KDD workflow applied in this study to extract insights from the available data. The first stage, data selection, involves choosing relevant failure data for the research objective. Next, data cleaning removes invalid or incomplete entries to ensure optimal data quality. Data modeling is then performed using the NB algorithm to identify patterns related to preventive maintenance needs and predict anticipated failure times.

Naïve Bayes Classifier

The NB Classifier is based on Bayes' Theorem, which provides a method for calculating the probability of a hypothesis given the available evidence. The basic formula for Bayes' Theorem is as follows:

$$\mathbf{P(H|E)} = \frac{P(H|E).P(H)}{P(E)} \quad (1)$$

Where:

- P(H|E) is the posterior probability, i.e., the probability of hypothesis H (e.g., class or label) given evidence E (e.g., data or attributes).
- P(E|H) is the likelihood of evidence E occurring given hypothesis H is true.
- P(H) is the prior probability of hypothesis H, representing our belief in the hypothesis before considering the evidence.
- P(E) is the marginal probability of evidence E, used to normalize the result.

The core mechanism of NB classification involves several key steps:

1. Calculating prior probabilities for each class label based on historical data.
2. Estimating the likelihood of each attribute value given the class label.
3. Applying Bayes' Theorem to compute the posterior probability for each class.
4. Choosing the class with the highest posterior probability as the prediction for each case.

Although this method assumes feature independence, it has proven effective in many classification applications, such as text recognition and spam filtering, due to its simplicity and computational speed.

Data Sources

The dataset employed in this research was derived from Daily Activity Reports (DAR) documented within PT ABC's Lakoni system. The reports encompassed failure incidents recorded over a six-month operational period across 47 units of Komatsu Dump Trucks HD785-7, resulting in 924 valid entries (data records). Each entry was categorized either as "Yes" (Breakdown Unscheduled, BUS, indicating the need for Preventive Maintenance) or "No" (Breakdown Scheduled, BS, classified as non-critical). This categorization facilitated a focused analysis of unscheduled failures, which represent a primary contributor to the observed decline in Physical Availability (PA).

Table 1. DAR Dataset sample after preprocessing

Machine ID	Breakdown Type	Problem Attribute	Shift	Label
DT7044	Tire	Low Power	Day	Yes
DT7052	Electrical System	Alternator Fault	Night	Yes
DT7071	Hydraulics	Brake Leak	Day	Yes
DT7086	Steering	Abnormal Noise	Day	No
DT7090	Electrical System	Starter Failure	Night	Yes

Data Preprocessing

Before the data is used for predictive modeling, preprocessing steps are performed to ensure optimal data quality. Invalid records and incomplete columns are removed from the dataset. Feature selection is then performed to choose relevant attributes for modeling, such as Machine ID, Breakdown Duration, Breakdown Type, Problem Attribute, and Shift. Categorical data is encoded to meet the requirements of the Naïve Bayes algorithm, while numeric variables are kept in their original form. This step is crucial to minimize noise that could affect the model's accuracy.

Modeling

The Naïve Bayes classification model was applied using RapidMiner Studio 9.6. The dataset was split into two parts: 75% for training (682 records) and 25% for testing (242 records). This data split ensures that the model does not overfit and can be tested on unseen data. To handle categories not found in the training data, Laplace smoothing is applied, ensuring model stability even when encountering unexpected categories in the test dataset.

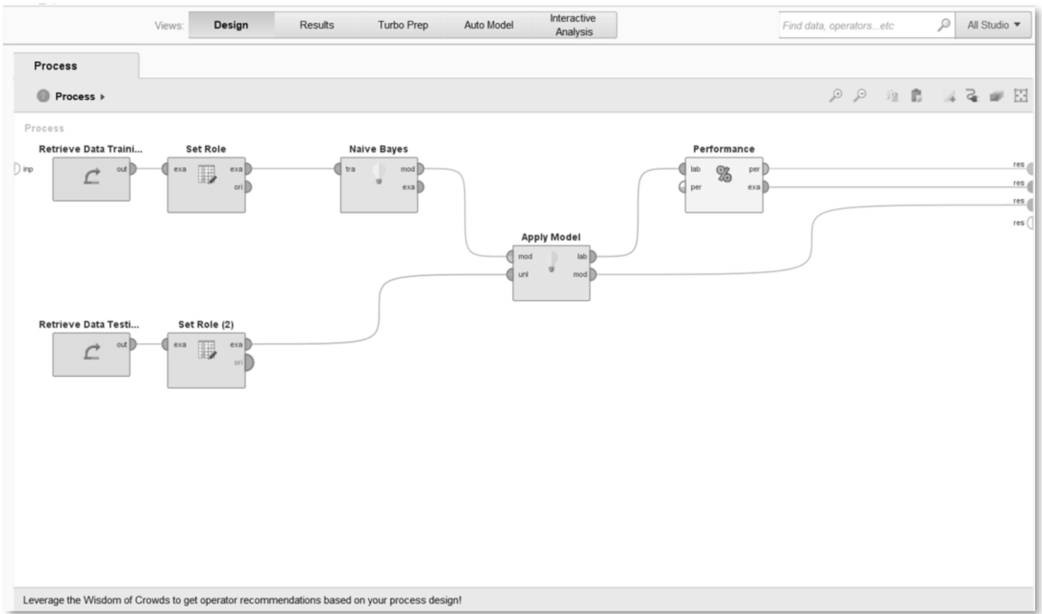

Figure 5. Naïve Bayes Data Modeling Process on RapidMiner Studio Platform

Evaluation Metrics

The performance of the model is evaluated using several standard metrics, including accuracy, precision, recall, and AUC (Area Under the Curve). Accuracy measures the proportion of correct predictions compared to the total number of predictions, while precision calculates the proportion of "Yes" predictions that correctly represent unscheduled failures. Recall indicates how many missed failures can be captured by the model. AUC is used to measure the model's ability to distinguish between both classes (scheduled and unscheduled failures) across various thresholds. Additionally, to ensure that the model does not overfit and can generalize well, 5-fold cross-validation is performed on the training data.

RESULTS AND DISCUSSION

Failure Analysis

A historical failure data analysis was conducted to identify dominant failure modes and validate the focus of the prescriptive model. The breakdown distribution revealed that a few subsystems were responsible for the majority of Dump Truck failures. As shown in Figure 6, tires were the most failure-prone component with 92 breakdown occurrences, followed by the electrical system with 68 occurrences. These two categories significantly outnumbered other components, such as the steering system, which recorded only a few incidents. This finding indicates that tires and electrical systems are the top priorities for reliability improvement, consistent with field observations where tires degrade rapidly in mining environments and electrical issues, such as wiring and alternator problems, occur frequently. Hence, prescriptive maintenance focusing on tire management (e.g., pressure checks and scheduled replacements) and electrical inspections is expected to deliver significant benefits.

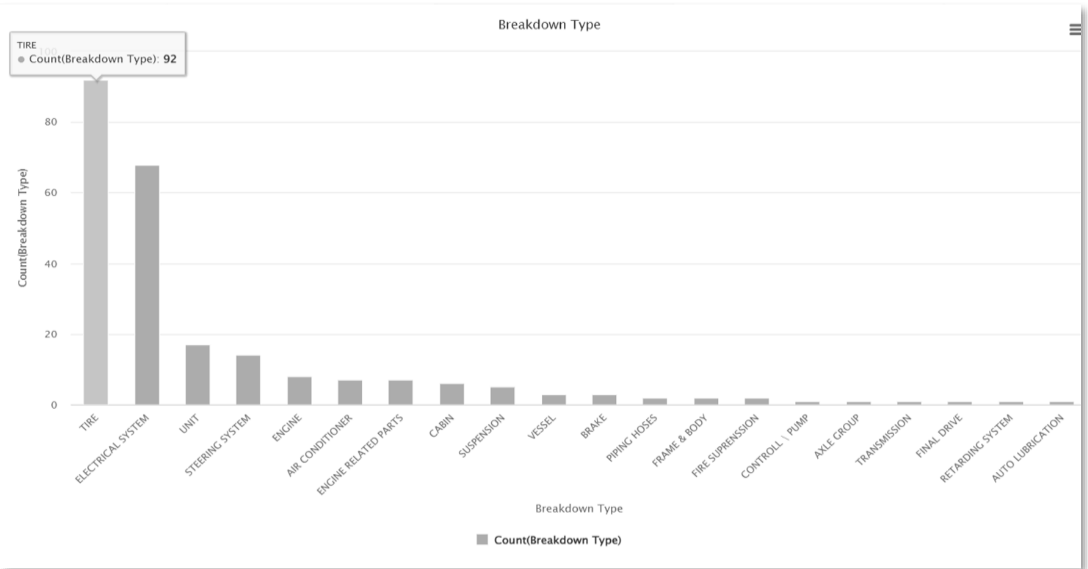

Figure 6. Breakdown Type

Furthermore, a detailed analysis of problem attributes (specific failure symptoms or anomalies) recorded in the Daily Activity Reports (DAR) was conducted. Although more than 100 failure descriptions were recorded, several issues recurred frequently. The most common was **"Low Power" (27 instances)**, typically indicating engine power loss, often related to fuel system or turbocharger problems. Other recurring problems included "Can't Start" and "Add Pressure." Identifying these patterns enables more targeted actions. For instance, if "Low Power" alerts often precede major engine failures, prescriptive actions such as cleaning fuel filters or inspecting the turbocharger should be implemented whenever these symptoms are detected in Figure 7.

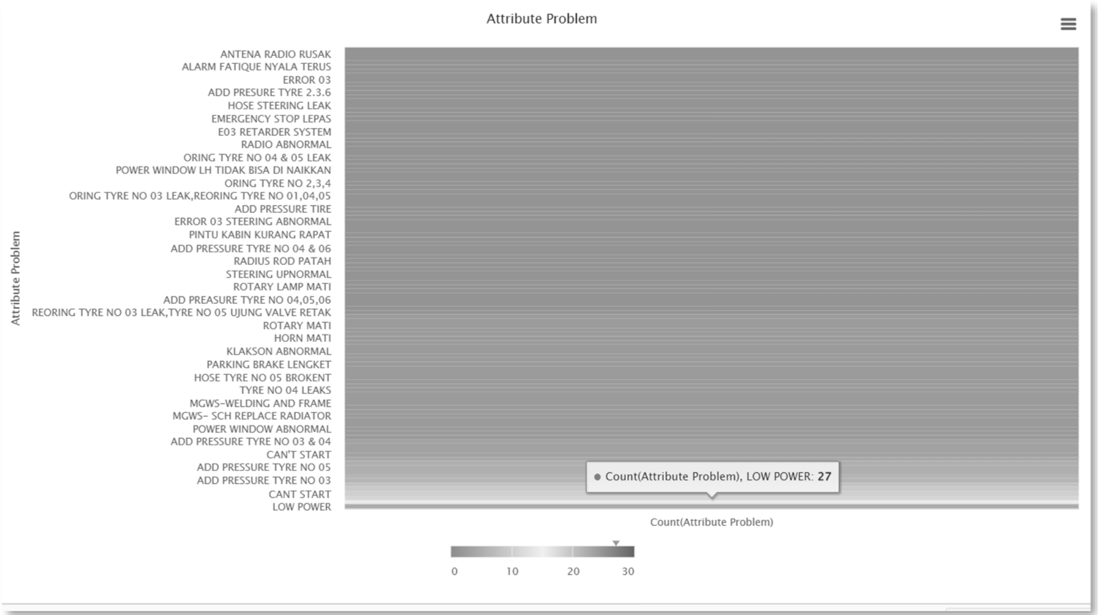

Figure 7. Attribute Problem

In addition, operational patterns also influenced failure risk. As illustrated in Figure 8, more failures occurred during Day Shifts (242 cases) compared to Night Shifts (93 cases). This may be attributed to heavier Dump Truck usage during the day or differences in crew activity. These findings suggest that daytime operations pose higher risks, requiring increased staffing or vigilance from maintenance crews. Such operational insights complement model predictions and can guide resource allocation in the field.

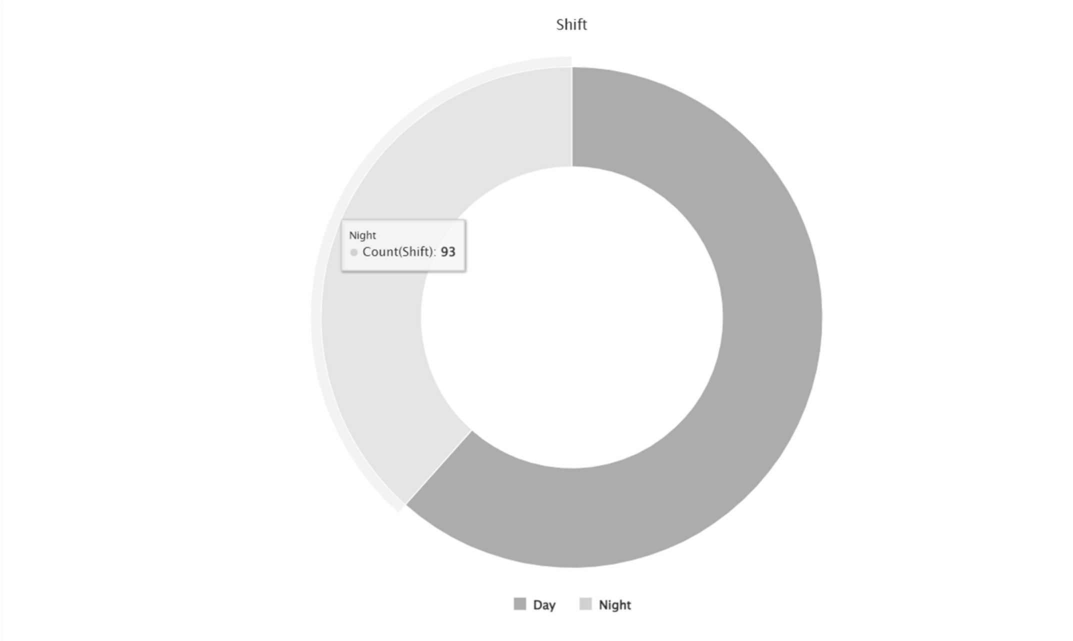

Figure 8. Work Shift

Based on the visual analysis and predictive modeling conducted using RapidMiner, unit DT7044 recorded the highest number of failures, with a total of 20 events, placing it in the high-risk category (Figure 9). This finding is consistent with the Naïve Bayes model results, which indicated that DT7044 has a very high likelihood of requiring preventive maintenance. The high frequency of failures in this unit reflects a decline in reliability and poses a potential threat to operational efficiency if not addressed promptly. Possible contributing factors include component fatigue, limitations in previous maintenance effectiveness, or specific operating conditions. Therefore, DT7044 should be prioritized in the prescriptive maintenance schedule through measures such as advanced inspections, scheduled component replacements, and shift-based monitoring. These actions will help reduce unscheduled downtime, control maintenance costs, and sustain equipment availability, while also reinforcing the implementation of RxM as a data-driven strategy in the field.

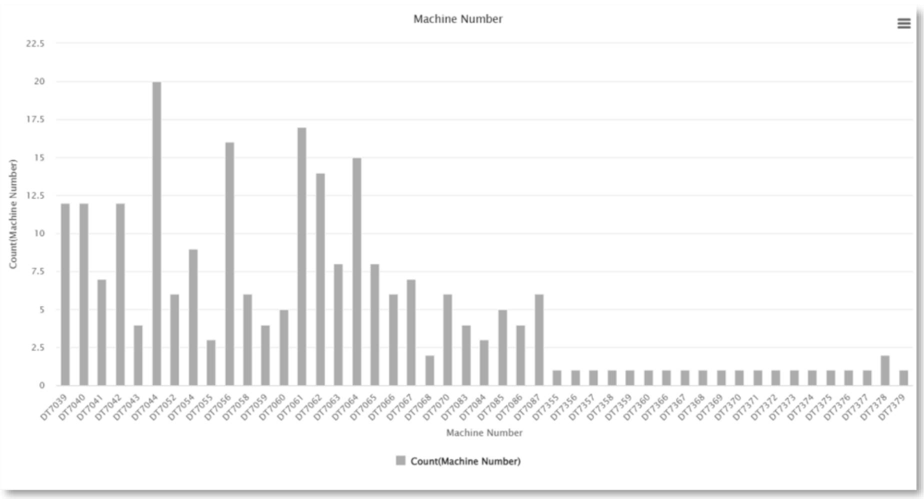

Figure 9. Machine number DT7044 has the highest number of failures

Model Performance

The Naïve Bayes model demonstrated excellent predictive performance on the test dataset (242 cases), achieving an accuracy of 97.93%, with 237 cases correctly classified. As summarized in Figure 9, the classification results were: 80 True Positives (TP), 157 True Negatives (TN), 0 False Positives (FP), and 5 False Negatives (FN). These results highlight the model's precision of 100% for the "Yes" class, recall of 94.12%, and accuracy of 97.93% overall.

Table View Plot View

accuracy: 97.93%

	true No	true Yes	class precision
pred. No	157	5	96.91%
pred. Yes	0	80	100.00%
class recall	100.00%	94.12%	

Figure 10. Naïve Bayes model performance results

The Performance Vector results (Figure 11) confirm these findings:

- Accuracy: 97.93% – Almost all predictions match the actual results, indicating that the model has a very high level of accuracy in classification.
- Precision (Yes): 100% – All "Yes" predictions provided by the model actually require maintenance, indicating that the model does not generate false alarms.
- Recall: 94.12% – Most units requiring maintenance were correctly detected, although there were some missed failures (false negatives).
- AUC: 0.995 – The model is very effective at distinguishing between cases that require maintenance ("Yes") and those that do not ("No") with near-perfect accuracy. This high AUC value indicates that the model has excellent discrimination capabilities shown in Figure 12.

Performance
Description
Annotations

PerformanceVector

```
PerformanceVector:
accuracy: 97.93%
ConfusionMatrix:
True:   No      Yes
No:     157     5
Yes:    0       80
precision: 100.00% (positive class: Yes)
ConfusionMatrix:
True:   No      Yes
No:     157     5
Yes:    0       80
recall: 94.12% (positive class: Yes)
ConfusionMatrix:
True:   No      Yes
No:     157     5
Yes:    0       80
AUC (optimistic): 0.995 (positive class: Yes)
AUC: 0.995 (positive class: Yes)
AUC (pessimistic): 0.995 (positive class: Yes)
```

Figure 11. Performance Vector

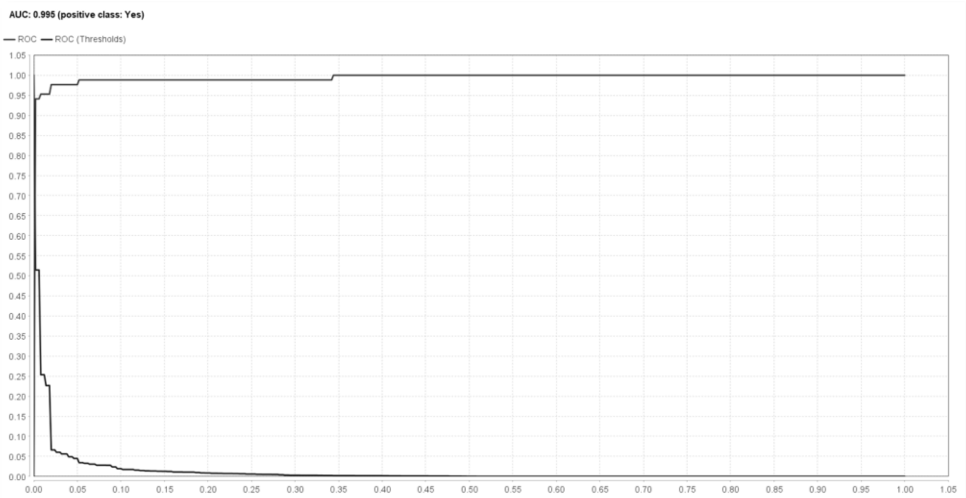

Figure 12. Area Under Curve (AUC)

Overall, these metrics confirm that the Naïve Bayes model is highly reliable for predicting maintenance needs, minimizing false positives while leaving room for improvement in reducing false negatives through model retraining and adjustment.

Implementation of Prescriptive Maintenance

The Naïve Bayes model was integrated into maintenance operations by updating procedures so that model outputs directly triggered actions. Figure 13 illustrates the new Standard Operating Procedure (SOP) integrating RxM into the maintenance workflow. The process consists of four key steps:

1. **Data Input**: Mechanics record breakdowns or anomalies in DAR forms, which are then verified by supervisors.
2. **Predictive Output**: Data is processed by the Naïve Bayes model, classifying cases as “Yes” or “No.” Site Technical Engineers (STE) review the outputs for accuracy.
3. **Maintenance Scheduling**: For confirmed “Yes” cases, planners issue preventive maintenance work orders, prioritizing the unit in the schedule.
4. **Feedback and Retraining**: Monthly reviews evaluate model performance, update datasets, and retrain the model when needed.

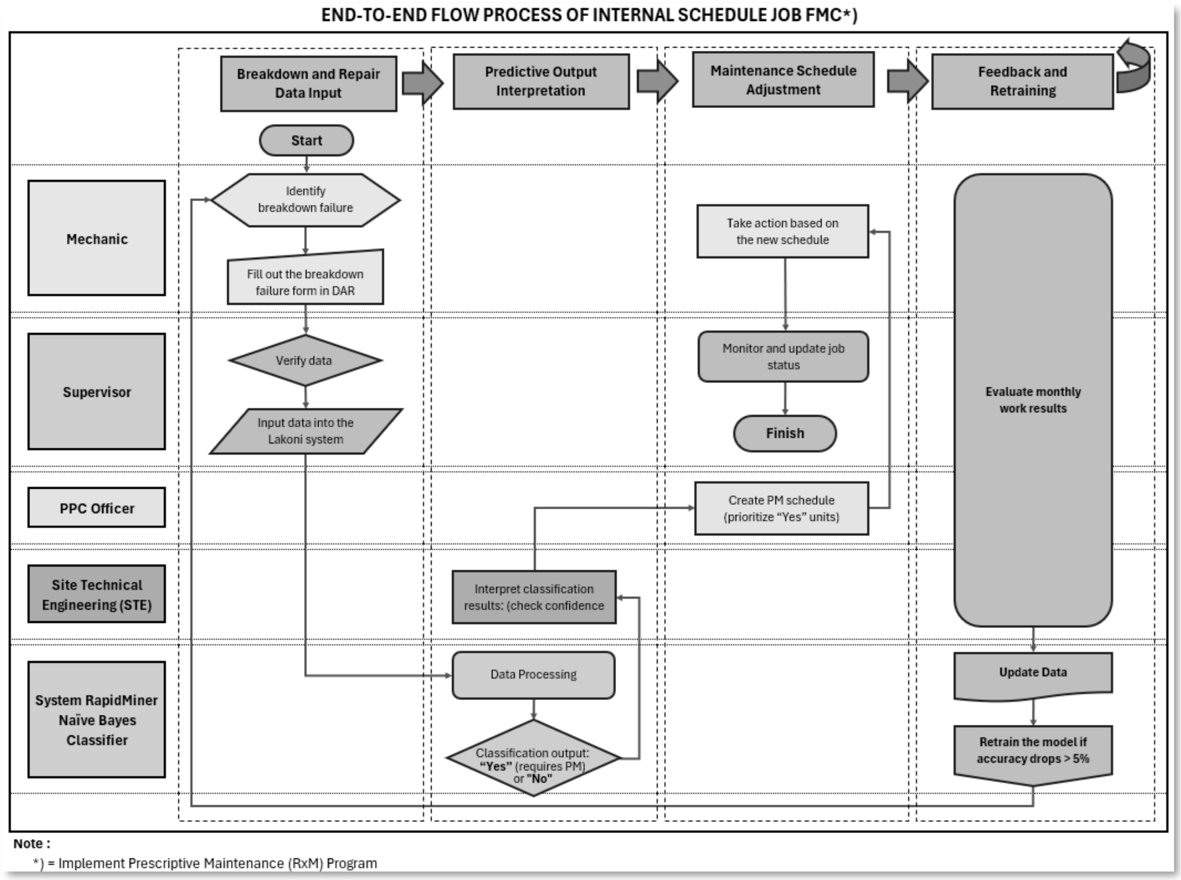

Figure 13. Standard Operating Procedure (SOP) for RxM Implementation

Key Performance Indicator Improvements

The implementation of RxM has led to substantial improvements in maintenance performance. As illustrated in Figures 14 and 15, the rate of BUS decreased significantly from 45% in 2024 to 26% in mid-2025, reflecting a shift of failures into planned maintenance events. This reduction directly contributed to higher PA, which improved steadily from 89.9% in January 2025 to 95.5% in June 2025, surpassing the contractual target of 92%.

The implementation of RxM has demonstrated tangible benefits, including reduced unplanned failures, improved scheduling, and optimized resource allocation. By transforming predictive insights into prescriptive actions, PT ABC achieved greater fleet reliability and productivity.

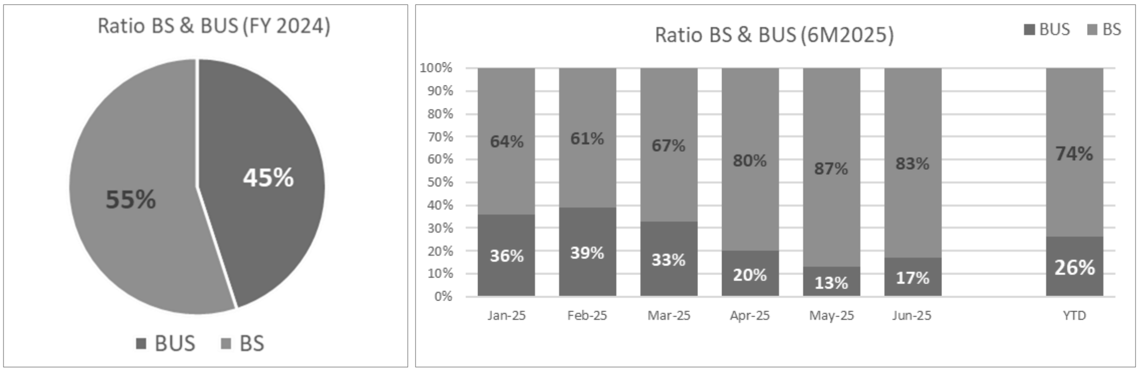

Figure 14. Standard Operating Procedure (SOP) for RxM Implementation

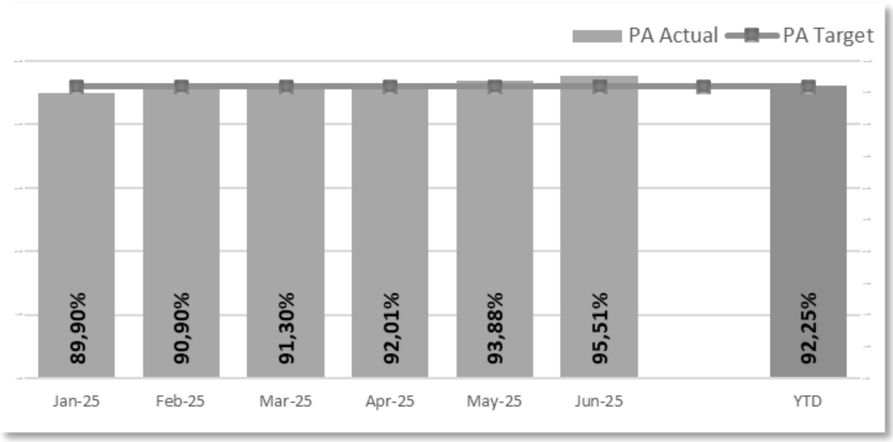

Figure 15. Physical Availability 2025

CONCLUSION

This study demonstrates that the implementation of RxM using the NB classifier can effectively address the research objective of improving the reliability and PA of Komatsu Dump Trucks HD785-7 under an FMC. By integrating the DMAIC framework with the KDD process, the research confirmed that historical failure data could be transformed into actionable insights for preventive maintenance scheduling. The predictive model produced highly reliable outcomes with 97.93% accuracy, 100% precision, 94.12% recall, and an AUC of 0.995, thereby validating the hypothesis that data-driven prescriptive analytics can significantly reduce BUS events and enhance operational performance. Furthermore, these results are consistent with previous studies, where Chazhoor et al. (2020) and Fajar et al. (2021) demonstrated the robustness of NB in predictive maintenance applications, while Moniri-Morad and Sattarvand (2023) highlighted the effectiveness of reliability modeling in identifying critical subsystems in mining dump trucks. Similarly, Rahimdel et al. (2024) employed a Bayesian network approach to strengthen reliability analysis in mining trucks, underscoring the value of probabilistic methods in maintenance decision-making. The alignment of this study's findings with prior research validates the proposed RxM framework and confirms its practical contribution to reducing BUS events and improving PA in heavy mining equipment fleets.

Beyond confirming its technical accuracy, the study highlights the strategic implication of RxM in shifting maintenance practices from reactive to proactive approaches, ultimately strengthening decision-making, optimizing resource allocation, and surpassing contractual PA targets. These discoveries indicate that prescriptive analytics can serve as a foundation for digital transformation in heavy equipment maintenance. Future research is encouraged to expand the framework by incorporating IoT-based telemetry, advanced machine learning algorithms, and cross-site applications to improve robustness and adaptability. Such developments will ensure that RxM continues to evolve as a sustainable, scalable solution for the mining industry and other heavy industrial sectors.

ACKNOWLEDGMENT

The authors would like to acknowledge PT ABC for providing access to operational data and facilities that supported the completion of this research. The constructive support and collaboration from the maintenance and operations team greatly contributed to the successful implementation of this study.

REFERENCES

Chazhoor, A., Mounika, Y., Sarobin, V. R., & Rao, V. N. (2020). Predictive maintenance using machine learning based classification models. *IOP Conference Series: Materials Science and Engineering, 954*(1), 012001. https://doi.org/10.1088/1757-899X/954/1/012001

Choubey, S., Benton, R. G., & Johnsten, T. (2021). A holistic end-to-end prescriptive maintenance framework. *Data-Enabled Discovery and Applications, 4*(1), Article 11. https://doi.org/10.1007/s41688-020-00045-z

Fajar, M., Yuliana, S., & Ramadhan, T. (2021). Implementasi Naïve Bayes Classifier untuk predictive maintenance mesin produksi. *Jurnal Sistem Informasi dan Teknologi Informasi, 10*(1), 23–31. https://doi.org/10.30630/jurti.v10i1.244

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI Magazine, 17(3), 37–54. https://doi.org/10.1609/aimag.v17i3.1230

Ferdinand, W. (2021). Predictive maintenance of heavy equipment in Indonesia leading heavy equipment company. *Journal of Mechanical Engineering Research, 14*(3), 92–98.

Giacotto, A., Marques, H. C., & Martinetti, A. (2023). Holistic and scalable smart optimization framework for prescriptive maintenance. In *Proceedings of ICAS 2024*. ICAS.

Limble. (2025, September 14). *The evolution of maintenance strategies*. Limble CMMS. https://limblecmms.com/learn/maintenance-operations/types-of-strategies/

Lumban Raja, H. D., Sunandar, M. A., Azlina, Y., Amrullah, A. M. K., Supriyono, S., Fauziningrum, E., & Kundori, K. (2024). Naïve Bayes classification and RapidMiner application for analysis of lecturer institution's performance. *AIP Conference Proceedings, 3065*(1), 030001. https://doi.org/10.1063/5.0232632

Mao, H., Liu, Z., Qiu, C., Huang, Y., & Tan, J. (2023). Prescriptive maintenance for complex products with digital twin considering production planning and resource constraints. *Measurement Science and Technology, 34*(12), 125903. https://doi.org/10.1088/1361-6501/aced5f

Moniri-Morad, A., & Sattarvand, J. (2023). A comparative study between the system reliability evaluation methods: case study of mining dump trucks. *Journal of Engineering and Applied Science, 70*, Article 103. https://doi.org/10.1186/s44147-023-00272-y

Pillai, R. S., Denny, P., & O'Connell, E. (2024). Optimizing predictive and prescriptive maintenance using unified namespace (UNS) for industrial equipment. *Preprints.org.* https://doi.org/10.20944/preprints202411.1037.v2

Rahimdel, M. J., et al. (2024). Bayesian network approach for reliability analysis of mining trucks. *Scientific Reports*, 14, Article 3415. https://doi.org/10.1038/s41598-024-52694-0