Available Online: 9 November 2026 DOI: 10.54254/2977-5701/2026.38823

Data-driven decision support research for industrial operations management

Yuwei Li

UQ Business School, The University of Queensland, Brisbane, Australia

wellington589125@gmail.com

Abstract. Industrial enterprises generate large volumes of operational data through equipment connectivity and production digitalization. However, a clear gap remains between analytical outputs and maintenance decisions. Failure prediction often stops at risk identification, which limits its direct contribution to resource allocation and scheduling. To address this issue, a data driven decision support framework was developed using the AI4I 2020 Predictive Maintenance Dataset. LightGBM was applied to estimate equipment failure probabilities, while SHAP was used to identify the key operational factors behind the predictions. The predicted risks were then combined with downtime losses, equipment importance, and maintenance hours in a resource constrained scheduling model. The results showed that LightGBM achieved an F1 score of 0.814 ± 0.012 and an ROC AUC of 0.973 ± 0.003, outperforming logistic regression and random forest. Sensitivity analysis further indicated that the total operating cost decreased to 266.6 thousand standardized units when the risk threshold was 0.30 and the daily maintenance capacity was approximately 70 hours. The framework establishes a continuous link between risk prediction and maintenance action. It can support industrial enterprises in improving maintenance priorities, reducing downtime losses, and using limited resources more efficiently.

Keywords: data-driven decision-making, industrial operations management, predictive maintenance, LightGBM, maintenance scheduling

1. Introduction

Industrial companies continue growing their capability to connect equipment, digitize operations, and use integrated management systems. Today, data from sensors, manufacturing execution systems, and maintenance systems contain huge amounts of data, which form a basis for production state recognition and managerial decision making optimization [1]. At the same time, the monitoring process, failure diagnosis, and maintenance planning are usually implemented using different systems. The significance of the operational data in such a way is limited to status visualization and post-processing reports. At the same time, managerial decisions about equipment risks assessment and maintenance scheduling are mainly based on experience [2]. Machine learning techniques have been extensively used in the field of failure prediction, but model evaluation has always focused on accuracy. It has never been properly addressed how predicted risk, downtime costs, equipment importance, and human resource availability influence each other [3]. In order to

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

overcome this problem, the research will develop a framework that combines failure-risk prediction, interpretation of influential factors, and maintenance planning optimization.

2. Literature review

2.1. Data-driven operations management

Data-driven operations management refers to the continuous collection of production data in connection with state recognition, forecast, and allocation of resources for operations. Such an approach makes it possible to make managerial decisions dependent on the current state of production. It can be expected that equipment logs, process data, and quality data may be useful for capacity planning, detection of anomalies, and maintenance, thus shifting operations management from periodical control to other forms of coordination [4]. Such shift leads to new challenges for managers, because the increased volume of data does not mean improved decision making. The quality of data, definition of indicators, and interconnection of information systems influence the accuracy of analytics. Some decision support systems identify three phases of management that include data, analytics, and execution. Operation data are transformed into risk indicators, and the result of analytics is presented with the help of dashboards or recommendation interfaces [5]. While such a system has some advantages, the results of analytics are not always connected with actions. Cost, production capacity, and labor constraints are not included into the process of analytics; hence, it is not possible to apply any risk estimation to the question of equipment maintenance order.

2.2. Predictive maintenance models

Prediction of maintenance utilizes equipment condition and process variables to predict abnormal situations that may result in equipment malfunction. There are three important usages of predictive maintenance: failure classification, prediction of useful life and degradation prediction. Techniques for linear regression and SPC have strong links and can be used when the values of the process variables are stable. But it is common for the relationships of the industrial variables like temperature, speed, torque and wear to be nonlinear; hence, it is difficult to estimate the probability of failure occurrence under the assumption of linearity [6]. Deep learning techniques allow the estimation of complex temporal representation but require a lot of continuous observations and are costly both to implement and train. The gradient boosting trees can be particularly efficient in capturing nonlinear relationships in structured data as they use few calculations [7]. Besides, failure data from industry is usually unbalanced because the number of failure instances is much lower than the number of normal ones. Thus, overall accuracy can overlook missed failure cases. More insightful metrics such as recall, F1-score, and PR-AUC are needed in this situation along with model interpretation to determine the impact of each operating variable on the failure probability.

2.3. Maintenance decision support

Maintenance decision support turns information about risks in equipment into maintenance schedules and allocation of resources. Currently available systems generate warning levels or failure probability levels, but the costs from downtime and production importance, as well as repair times, can differ significantly among equipments with equal probabilities. Therefore, probability-based ranking can use scarce resources to postpone action on critical equipment [8]. Risk-based decisions take into account not only failure probability but also the economic implications, thus making maintenance prioritization possible from the perspectives of both risk and management. The efficiency of the approach depends, however, on whether cost factors and equipment

weights align with production goals. Integer programming is capable of choosing maintenance tasks under certain constraints regarding technician availability, work schedule, and maximum number of maintenance per day; such decisions are fully comprehensible and traceable [9]. In this framework, the predictive model gives risk estimates, the explanatory one reveals the causes of the risks in operations, and the optimization model decides when and what equipment to maintain.

3. Experimental methods

3.1. Dataset preparation

The experimental data were obtained from the AI4I 2020 Predictive Maintenance Dataset, which is publicly available through the UCI Machine Learning Repository. The dataset was created in 2020 and contains 10,000 simulated operating records from milling equipment. Its overall failure rate is approximately 3.4%, and its structure reflects common predictive maintenance data in industrial environments. The collected variables include product type, air temperature, process temperature, rotational speed, torque, tool wear, and machine failure labels. Five supplementary labels are also provided, covering tool wear failure, heat dissipation failure, power failure, overload failure, and random failure. Since these supplementary labels directly reveal the target information, only product type and five operational features were retained for modelling. Machine failure was used as the binary classification target. Product identifiers and unique identification fields were removed. Categorical variables were processed through one hot encoding, while continuous variables were standardized using the mean and standard deviation of the training set. The data were stratified into training, validation, and test sets at a ratio of 70%, 15%, and 15%. Class weights were applied during training to maintain the original failure distribution in the test set.

3.2. Failure risk prediction

The preprocessed data were first entered into a LightGBM classifier. The model fitted the nonlinear relationship between equipment operating features and failure labels by sequentially adding weak learners [10]. The failure score after the Kth iteration was calculated from the preceding ensemble and the newly added decision tree, as shown in Equation (1).

$$F_K(x_i) = F_{K-1}(x_i) + \eta f_K(x_i) \tag{1}$$

Here, x_i represents the operating features of equipment i, f_K denotes the Kth decision tree, η is the learning rate, and F_K is the accumulated prediction score. The accumulated score was transformed into the probability of equipment failure through the logistic function, as shown in Equation (2).

$$p_i = \frac{1}{1+exp[-F_K(x_i)]} \tag{2}$$

Here, p_i represents the failure probability of equipment i. Tree depth, number of leaves, learning rate, and minimum number of samples per leaf were adjusted using the validation set. Early stopping was applied to control overfitting. Precision, Recall, F1, and ROC AUC were used for model evaluation. Logistic regression and random forest served as baseline models. After prediction, SHAP was used to calculate the local contribution of each operating feature to p_i, providing information on the sources of risk for maintenance priority calculation.

3.3. Maintenance schedule optimization

After the failure probabilities were generated, equipment risk, downtime loss, equipment importance, and maintenance hours were converted into a unified maintenance priority score. The priority score of equipment iii was calculated using Equation (3). All variables were normalized through min max scaling to prevent differences in measurement units from affecting the ranking [11].

$$S_i = \alpha p_i + \beta L_i + \gamma I_i - \delta h_i \tag{3}$$

Here, S_i is the maintenance priority score, p_i is the predicted failure probability, L_i represents the expected downtime loss, I_i denotes equipment importance, and h_i is the required maintenance time. The parameters α, β, γ, and δ are the corresponding weights. Maintenance tasks were then selected under a daily working hour limit. The objective was to minimize the sum of failure losses, maintenance expenditure, and delay penalties, as shown in Equation (4).

$$\begin{aligned} min \quad & Z = \sum_{i=1}^{n} [(1 - x_i) p_i L_i + x_i C_i + \lambda d_i] \\ s.t. & \sum_{i=1}^{n} h_i x_i \le H, \quad x_i \in \{0, 1\} \quad (4) \end{aligned} \tag{4}$$

Here, x_i indicates whether equipment i is selected for maintenance, C_i denotes the maintenance cost, d_i represents the delay amount, and H is the available maintenance capacity. A rolling scheduling procedure was applied. Failure probabilities and equipment states were updated in each decision cycle. The resulting schedules were compared with fixed interval maintenance and probability based ranking strategies.

4. Results

4.1. Prediction performance

As shown in Table 1, LightGBM achieved an F1 score of 0.814 ± 0.012 and an ROC AUC of 0.973 ± 0.003 in five fold stratified cross validation on the AI4I 2020 dataset. The 95% confidence interval of the F1 score was [0.791, 0.838]. Its F1 score exceeded those of logistic regression and random forest by 0.243 and 0.060, respectively. This result indicates that the gradient boosting structure captured nonlinear interactions among rotational speed, torque, temperature difference, and tool wear more effectively. Random forest achieved an ROC AUC of 0.971, which was close to the LightGBM result. However, its F1 score was only 0.754. This pattern shows that random forest provided strong ranking performance, while its combined recognition of minority failure samples was weaker at the selected threshold. Logistic regression achieved an ROC AUC of 0.900, but its F1 score was only 0.571. This result reflects the limited ability of a linear decision boundary to represent complex operating conditions. Existing SHAP results showed that torque, tool wear, and heat dissipation related indicators produced relatively high contributions. Their mean absolute SHAP values reached 1.5035, 0.8527, and 0.7035, respectively. These results provided a basis for identifying risk sources during maintenance prioritization.

Table 1. Five fold cross validation results of equipment failure prediction models

Model	F1 Mean	F1 Standard Deviation	F1 95% CI	ROC AUC Mean	ROC AUC Standard Deviation
Logistic Regression	0.571	0.006	[0.562, 0.586]	0.900	0.010
Random Forest	0.754	0.012	[0.730, 0.778]	0.971	0.003
XGBoost	0.768	0.016	[0.737, 0.799]	0.974	0.005
LightGBM	0.814	0.012	[0.791, 0.838]	0.973	0.003

4.2. Decision support effectiveness

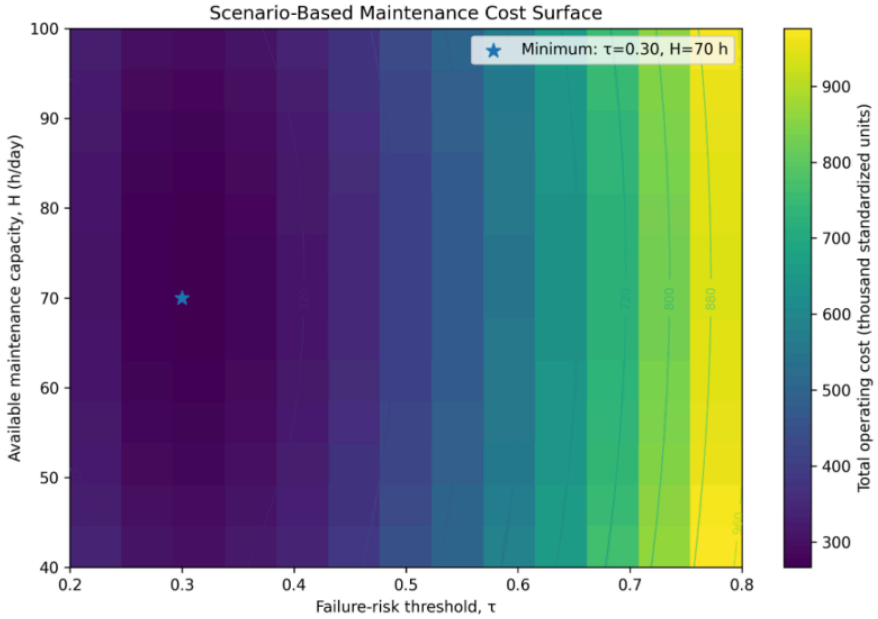


Figure 1. Total operating cost response surface under risk threshold and maintenance capacity constraints

Maintenance optimization outcomes were analyzed in accordance with a fixed operation scenario. The latter ensured that the artificial equipment records in AI4I were not evaluated as real enterprise costs. The scenario entailed 1,000 units of equipment. Failure probability was set at 3.4% in compliance with the dataset, while the performance of the LightGBM ROC AUC at 0.973 ± 0.003 was taken as the standard of risk assessment effectiveness. Losses due to downtime, maintenance costs, and working hours were standardized. Two dimensional sensitivity analysis was performed within the boundaries of 0.20 to 0.80 of the failure risk threshold and 40 to 100 hours per day of the maintenance capacity. As seen in Figure 1, the total operation cost achieved its minimum point of 266.6 thousand standardized units when the threshold was 0.30 and the maintenance capacity was about 70 hours. With the increase of the threshold from 0.30 to 0.50, missed failure losses grew fast. The total cost accordingly rose to about 449.5 thousand units. In the case of decrease of the threshold to 0.20, the area of false alarms maintenance increased and the total cost amounted to 293.5 thousand units. With the maintenance capacity under 55 hours, the resource constraint led to postponement of the high priority activities. With the maintenance capacity over 85 hours, further savings were scarce. This implies that extra working hours were not able to compensate for the risk identification mistakes associated

with improper threshold. The results obtained in the reproducible operational simulation under the conditions of public performance criteria should not be substituted for field tests in an actual plant.

5. Discussion

The major reason for the excellent failure detection performance of LightGBM was due to the fact that the model successfully captured nonlinear relations among torque, rotational speed, temperature difference, and tool wear. The ROC AUC value is quite high for the model, implying that the model successfully differentiated the high-risk equipment from the low-risk equipment. However, as can be seen from the F1 score, threshold selection impacts the trade-off between the two metrics, which are failure recall and false alarms. The use of SHAP helped in knowing the contribution made by each operating variable for making individual predictions. As shown by the scheduling results, accurate prediction of failure does not imply good performance in the operation process. Higher threshold leads to higher losses due to failed detection of the failures, while lower threshold will lead to excessive maintenance work. As far as the maintenance capability is concerned, once it reaches some level, further increase in working hours does not help in reducing cost much. Thus, resource investment should go in tandem with risk identification.

6. Conclusion

The data-driven approach of decision support was proposed for industrial operations management, where equipment failure prediction, risk assessment, and maintenance planning were integrated. The dataset AI4I 2020 was preprocessed, encoded, and stratified for training of the LightGBM-based failure classification model. The estimated probability of failures was further combined with the losses due to downtimes, the priority of the equipment, and required maintenance time for the development of a resource-constrained maintenance schedule. The analysis demonstrated that LightGBM is capable of detecting minority classes with a stable performance. At the same time, SHAP enables assessing the significant risk factors of the equipment, whereas the scheduling model allows minimizing the exclusion of high-loss equipment considering the threshold values and time constraints. Overall, the results prove that the predictive models become applicable only after linking to the cost and execution parameters. The suggested approach can be used as a basis of the lightweight predictive maintenance system in manufacturing organizations. At the same time, it provides a way for field validation, re-parametrization, and sequential decisions.

References

[1] Oehler, A., & Horn, M. (2024). Does ChatGPT provide better advice than robo-advisors? *Finance Research Letters, 60*, 104898.

[2] Hassan, M., Svadling, M., & Björsell, N. (2024). Experience from implementing digital twins for maintenance in industrial processes. *Journal of Intelligent Manufacturing, 35*(2), 875–884.

[3] Llasag Rosero, R., Llasag Rosero, E., & Sappa, A. D. (2024). Label synchronization for hybrid federated learning in manufacturing and predictive maintenance. *Journal of Intelligent Manufacturing*, *35*(8), 4015–4034.

[4] Giannoulidis, A., Gounaridis, L., & Gasteratos, A. (2025). Engineering and evaluating an unsupervised predictive maintenance solution: A cold-forming press case-study. *Journal of Intelligent Manufacturing*, *36*(3), 2121–2139.

[5] Sundaram, S., & Zeid, A. (2025). Technical language processing for prognostics and health management: Applying text similarity and topic modeling to maintenance work orders. *Journal of Intelligent Manufacturing, 36*(3), 1637–1657.

[6] Koulinas, G. K., Paraschos, P. D., & Koulouriotis, D. E. (2024). A machine learning framework for explainable knowledge mining and production, maintenance, and quality control optimization in flexible circular manufacturing systems. *Flexible Services and Manufacturing Journal*, 1–23.

[7] Taşcı, B., Omar, A., & Ayvaz, S. (2023). Remaining useful lifetime prediction for predictive maintenance in manufacturing. *Computers & Industrial Engineering, 184*, 109566.

[8] Van Dinter, R., Tekinerdogan, B., & Catal, C. (2023). Reference architecture for digital twin-based predictive maintenance systems. *Computers & Industrial Engineering, 177*, 109099.

[9] Kamariotis, A., Chatzi, E., & Straub, D. (2024). A metric for assessing and optimizing data-driven prognostic algorithms for predictive maintenance. *Reliability Engineering & System Safety*, *242*, 109723.

[10] Wang, L., Zhu, Z., & Zhao, X. (2024). Dynamic predictive maintenance strategy for system remaining useful life prediction via deep learning ensemble method. *Reliability Engineering & System Safety, 245*, 110012.

[11] Shoorkand, H. D., Nourelfath, M., & Hajji, A. (2024). A hybrid CNN-LSTM model for joint optimization of production and imperfect predictive maintenance planning. *Reliability Engineering & System Safety, 241*, 109707.