



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Recommendation of a machine learning model

Student Name: Suhaana Kapali

Student ID: 2538421

MSc Data and Artificial Intelligence

Module: L7- Application of Machine Learning

Module Code: ACCA7025

Lecturer: Dr. Aliyu Sambo

Word Count: 3369

Table of Contents

1. Introduction	3
1.1 Objectives	4
2. Literature Review	4
2.1 Predictive maintenance with traditional machine learning algorithms	4
2.2 Performance evaluation of predictive maintenance	6
2.3 Major Challenges and Solutions	6
2.3.1 Computational Cost	6
2.4 Implications for the present study	7
3. Proposed Workflow Methodology	8
3.1 Overall Workflow	8
3.2 Data Preparation	9
3.2.1 Data Collection	9
3.2.2 Data Preprocessing	9
3.2.3 Feature Engineering	9
3.2.4 Addressing Class Imbalance	10
3.2.5 Data Splitting	10
3.3 Model Selection	11
3.4 Model Evaluation Metrics	13
3.4.1 Primary Metrics	13
3.4.2 Cross-Validation for Effective Evaluation	14
3.5 Project Timeline	14
4. Discussion and Recommendation	15
4.1 Algorithm Comparison and Critical Analysis	15
4.1.1. The inadequacy of simple linear and distance-based models:	15
4.1.2. The Performance vs. Interpretability Trade-off of complex models:	15
4.1.3. The Superior Balance of Random Forest:	16
4.2 Recommendation: Random Forest	17
4.3 Evaluation Recommendations	18
5. Conclusion	19
Bibliography	20

Abstract

This report presents a consultancy recommendation for a manufacturing firm that wants to implement a predictive maintenance solution using traditional machine learning under limited computational resources. The main task is to select an algorithm that balances predictive accuracy, model interpretability, and computational cost. A review of recent literature is conducted to critically review five candidate algorithms that are logistic regression, K-nearest neighbors, support vector machine, XGBoost, and random forest in this report. The evaluation emphasizes their suitability for dealing with complex, non-linear sensor data like temperature, vibration, and pressure; their robustness to class imbalance and their real world feasibility in view of the limitations of the company. The analysis ends up with Random Forest as the better option. The ensemble methodology is highly accurate and robust; the essential feature importance ranking provides important interpretability for operational insights, and its moderate computational requirements guarantee possibility. This recommendation is supported by a complete workflow methodology and a proposed evaluation framework based on the precision, recall, and F1-score, which are important metrics for imbalanced failure prediction.

1. Introduction

One of the biggest operational and financial risks in today's manufacturing is unplanned equipment shutdown. According to industry research, predictive maintenance has the potential to cut downtime by 30% to 50%, lower maintenance expenses by 12% to 18%, and delay equipment duration by 20% to 40% [1]. Due to the limited computational resources of the manufacturing company, a realistic approach is required in the implementation of machine learning, without resorting to computationally expensive deep learning methods. The company collects real sensor data like temperature, vibration, and pressure from the machinery in operation.

This data acts as a key tool in predicting failures before they occur, thus allowing for planned maintenance scheduling instead of an emergency reaction. The only problem is

how to choose a suitable machine learning method to efficiently handle this data processing with limited resources.

1.1 Objectives

- To conduct a methodical evaluation of the pros and cons of the candidate algorithms, which are Logistic Regression, K-NN, SVM, XGBoost, and Random Forest, controlled by industrial settings.
- To suggest a systematic approach to build and assess a predictive maintenance model, such that it includes data preprocessing and performance measurement.
- Recommending a single algorithm achieved a good balance among accuracy, interpretability, and computational expense in a justified manner.
- To develop appropriate and industry specific evaluation metrics like F1-score, precision, and recall for the suggested model.

2. Literature Review

Predictive maintenance represents a strategic shift from reactive and preventive maintenance approaches. Predictive maintenance is about using data driven techniques to predict when something will break and maintain it at the best time. It is not about fixing things after they break or performing maintenance at scheduled periods [2]. This model is made possible by the production of sensors in industrial machines and advances in machine learning techniques.

The basic PdM method is based on a systematic pipeline that is sensor data possession, feature extraction, irregularity identification, problem diagnosis, and remaining usable life calculation. This procedure provides the basis for the implementation of predictive maintenance solutions based on machine learning.

2.1 Predictive maintenance with traditional machine learning algorithms

Recently, various machine learning methods were systematically reviewed for predictive maintenance purposes. A detailed assessment with the NASA CMAPSS dataset to compare ten traditional machine learning algorithms and three deep learning models in binary failure prediction has been done [3]. The gradient boosted tree models turned out

to be highly efficient alternatives to resource heavy deep learning models and were also quite competitive in effectiveness among the standard methods. There are various algorithms that are often suggested in literature as having great potential for predictive maintenance, such as logistic regression, which is a classic linear classifier, and it is highly interpretable and computationally cheap [4].

But due to the modeling of linear relationships between variables and log odds, its performance on complex nonlinear machine data is quite limited. One study has shown that this technique may reach the maximum accuracy level, which is an indication of its weak capability of detecting complicated failure patterns.

K-Nearest Neighbors is a distance based classification method and is known for its simplicity and non parametric character. Studies reported an accuracy of 89.13% in certain applications. However, its sensitivity to feature scaling and the computing cost per forecast delay limit large scale application. KNN can provide immediate predictions following preprocessing, which is attractive to time critical applications [5].

The Support Vector Machine maximizes the margin and works well in high dimensional spaces using kernel transformations. It is well known that SVM performs well with some studies reporting good test accuracy [6]. However, there are tradeoffs in processing requirements and poorer interpretability in comparison to tree based models.

The Random Forest Classifier is a reserved cooperative of decision trees, which is strong to overfitting and provides great performance on imbalanced datasets such as those found in maintenance scenarios. New research supports the usefulness of RFC, with some studies reporting 100% accuracy in some fault detection situations [7]. Random Forest also provides a rating of feature relevance, which helps the interpretability of the model. XGBoost uses gradient boosting combined with highly advanced regulation techniques. It has been at the top of performance competition with other methods in PdM, and its accuracy has been reported as being as high as 98.3% But it is capable of parallel processing and is at the peak as far as combination learning is concerned [8].

It represents a very good choice, but to train it, more money needs to be spent than using simple models.

2.2 Performance evaluation of predictive maintenance

Accuracy is a traditional performance measure, but it is often unreliable in imbalanced data sets where failures are a small percentage of observations. A model that always predicts 'no failure' would have high accuracy but be operationally useless. Therefore, precision, recall, and F1-score are more important to evaluate the model's ability to detect rare failure cases. However, there is an essential tension between these metrics, and a compromise must be made between the business cost of false alarms and much higher costs of undetected failures. [9].

2.3 Major Challenges and Solutions

One of the major problems when dealing with maintenance datasets is data imbalance, as the datasets are usually very large and failure records are only a small portion of the whole set. One of the ways that the problem of data imbalance can be dealt with is by applying the Synthetic Minority Oversampling Technique SMOTE, which increases the model sensitivity of minority classes. Industrial implementation of complicated models is difficult because the models are typically black boxes. Several studies have made use of SHAP values to explain model predictions and to explain them globally in response to the questions of what and why.

2.3.1 Computational Cost

Real time PdM needs very fast model reasoning running on cheap hardware. Deep learning models can be very accurate but require a lot of computing power and need a GPU to run effectively. This not only raises the price of the deployment but also excludes smaller and medium sized factory units, which usually have very limited IT budgets. As a result, Random Forest is chosen as the ultimate classifier for its contribution of high accuracy, strength, and computational efficiency and all three features that allow it to

make predictions in a matter of milliseconds on ordinary hardware without the presence of GPUs [10]. This lets the whole thing be cheap and practical to use in real settings of industrial environments. The justification of picking Random Forest instead of more advanced algorithms is supported further by quantitative data on power consumption. [11] This proved that Random Forest's energy usage during training is significantly lower than that of Support Vector Machines despite their similar accuracy, which is 97.11% and 97.65%. That's why Random Forest is certainly the greener option, plus a more pocket friendly one for deployment of usual hardware in an industrial setting where energy and CPU power are major limiting factors.

2.4 Implications for the present study

The literature review indicates that there is no one algorithm that can be considered the best one in all situations. The best solution relies on what values the most. Like, is it accuracy, interpretability, computational efficiency, or business cost? In this case the algorithm must balance all three parameters while working under resource restrictions. Random Forest is a good competitor, as it has good performance, allows us to evaluate feature importance, and has manageable computational requirements.

3. Proposed Workflow Methodology

3.1 Overall Workflow

The proposed workflow builds upon a structured machine learning pipeline adapted from established PdM frameworks.

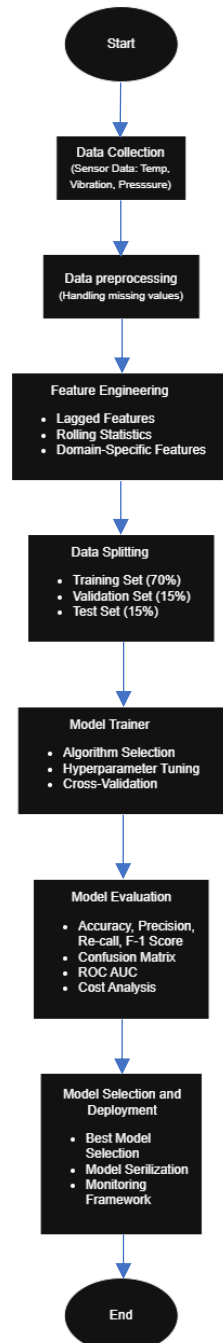


Figure 1 Structured machine learning pipeline

3.2 Data Preparation

3.2.1 Data Collection

The company collects sensor data such as temperature, vibration, and pressure. Additional features such as rotational speed and force are also a great predictor of failure based on literature analysis. The data must be time marked so that time series analysis and feature engineering can be done.

3.2.2 Data Preprocessing

Missing Value Treatment:

Due to communication errors or sensor failure, the sensor data may have missing values. Short gaps are filled by sudden occurrence. Extended missing periods are deleted.

Outlier Detection: Unusual readings should be detected by statistical methods, for example, Z-score or IQR based detection. Outliers should be explored as a possible failure indicator rather than automatically removed.

Normalization/Scaling: Normalize features so that all variables are on the same scale. Scale the features to the range, which is 0, 1, using min or max scaling for distance based algorithms. For scale invariant tree based methods, standardization with mean=0 and std=1 is suitable.

3.2.3 Feature Engineering

Features should be based on known PdM practices such as:

Features based on time:

- Lagged values for each sensor
- Rolling statistics like mean, standard deviation, minimum, and maximum over sliding windows

Rate of change characteristics:

- First derivative that changes between successive readings
- Acceleration, which is the second derivative

Domain-specific features:

- Product type indicators
- Running hour accumulators

Predictive accuracy improves significantly when using the entire feature set. Research shows that feature engineering improves model performance considerably.

3.2.4 Addressing Class Imbalance

Because equipment failures are rare, the problem of class imbalance must be tackled. Due to the lack of failures, the dataset will be naturally class imbalanced. The Synthetic Minority Oversampling Technique SMOTE will be used to solve this problem. SMOTE synthetically creates examples of the minority failure class to avoid the model being biased towards the dominant “normal” class. This technique will be used during the data preparation phase before the model training. Other approaches are the following:

- Under sampling: Reducing the samples of the majority class
- Oversampling: Repeating Minority Class Samples
- Cost-sensitive learning: Making misclassification costs higher for failure cases

3.2.5 Data Splitting

The dataset will be split into training and testing sets with an 80/20 stratified split. This split will help us to not only carry out a detailed model evaluation but also to avoid data leakage. Given the very low number of instances of the failure class, it's important to plan so that both train and test sets have the same failure ratio. The model is first trained on the training set, then hyperparameters are tuned and cross-validated. The test set is then only used for the final model evaluation.

3.3 Model Selection

A literature review was performed, and considering the different criteria for this predictive maintenance application, i.e., the need to resolve predictive accuracy, model interpretability, and computational cost, five traditional machine learning algorithms were selected for evaluation which is Logistic Regression, K-Nearest Neighbors KNN, Support Vector Machine SVM, XGBoost, and Random Forest.

The justification for choosing these algorithms is their range of complexity and applicability to predictive maintenance tasks:

- Logistic Regression is a simple linear model that is interpretable and basic.
- KNN is an example of distance based instance learning which is fast in prediction.
- Support Vector Machine SVM can perform powerful non-linear classifications by way of kernel transformations.
- XGBoost has great performance via gradient boosting.

The Random Forest is a combination method that has essential interpretability.

Based on literature analysis, three candidate algorithms are recommended for evaluation:

Algorithms	Strengths	Weaknesses	Suitability
Random Forest	Effective to overfitting, Handles non-linearity, Provides feature importance for interpretability, <u>Good</u> performance on imbalance data.	Limited Interpretability compared to single trees, average training time.	High: Best balance of accuracy, interpretability, and computational cost.
XG Boost	Competitive accuracy, Regularization, Parallel processing	Less interpretable, Higher training cost, More complicated tuning	Medium: Highest accuracy but computational transparency.
Logistic regression	High interpretable, Very fast training, Simple implementation	Linear assumptions, Limited performance on complicated data	Low: Good model but likely not enough accuracy.
Support Vector Machine	Effective in high dimensional spaces, Handles nonlinearity through kernel transformations, Strong performance on complicated data.	High computational cost during training, Poor interpretability, sensitive to parameter tuning, Memory demanding.	Medium: Strong performance but resource demanding and lacks interpretability.
K-Nearest Neighbors	Simple and intuitive, No specific training phase, very fast predictions, Effective with smaller datasets.	Performance degrades with high dimensional data, Sensitive to feature scaling, No essential feature importance.	Medium to low: Good for rapid deployment but less suitable for complicated, large scale industrial data.

3.4 Model Evaluation Metrics

3.4.1 Primary Metrics

In predictive maintenance, the cost of missing a failure is much higher than the cost of a false alarm, so choosing the right evaluation metrics is critical. The standard metrics such as accuracy can be very misleading in equipment failure data because there is an essential class imbalance where failures typically include a very small percentage of all operational readings. A model could get 99% accuracy by just predicting 'no failure' every time but can be useless operationally [12]. The following metrics are therefore recommended for comprehensive model evaluation.

Metric	Description	Importance in Predictive Maintenance
Precision	Proportion of predicted failures that are actual failures.	Minimizes false alarms, saving maintenance resources and maintaining operator trust.
Recall	Proportion of actual failures successfully detected.	Highest Priority: Minimizes the costliest missed failures i.e. false negatives
F-1 score	Harmonic mean of Precision and recall.	Main Evaluation Metric: Provides a single and balanced metric for <u>models</u> comparison of imbalanced datasets.
Accuracy	Overall proportion of correct predictions.	Can be misleading when failures are rare.
ROC-AUC	Discriminative ability across all threshold.	Useful for model comparison but will be supplemented with PR-AUC as failures are rare.

3.4.2 Cross-Validation for Effective Evaluation

To ensure the reliability of the evaluation, each model will be tested with a stratified 5-fold cross-validation. The average and the standard deviation of the F1-score, precision, and recall will be reported over all folds. This approach provides a stronger measure of how well the model works than a single train test split and makes it less likely that an especially good or bad split will occur.

The best hyperparameters will be selected using cross-validation, and the final model will be trained on the entire training dataset using these hyperparameters. The final model will be evaluated on the reserved test dataset to evaluate the generalization performance to new data.

3.5 Project Timeline

The following Gantt chart illustrates the proposed implementation timeline for the predictive maintenance project:

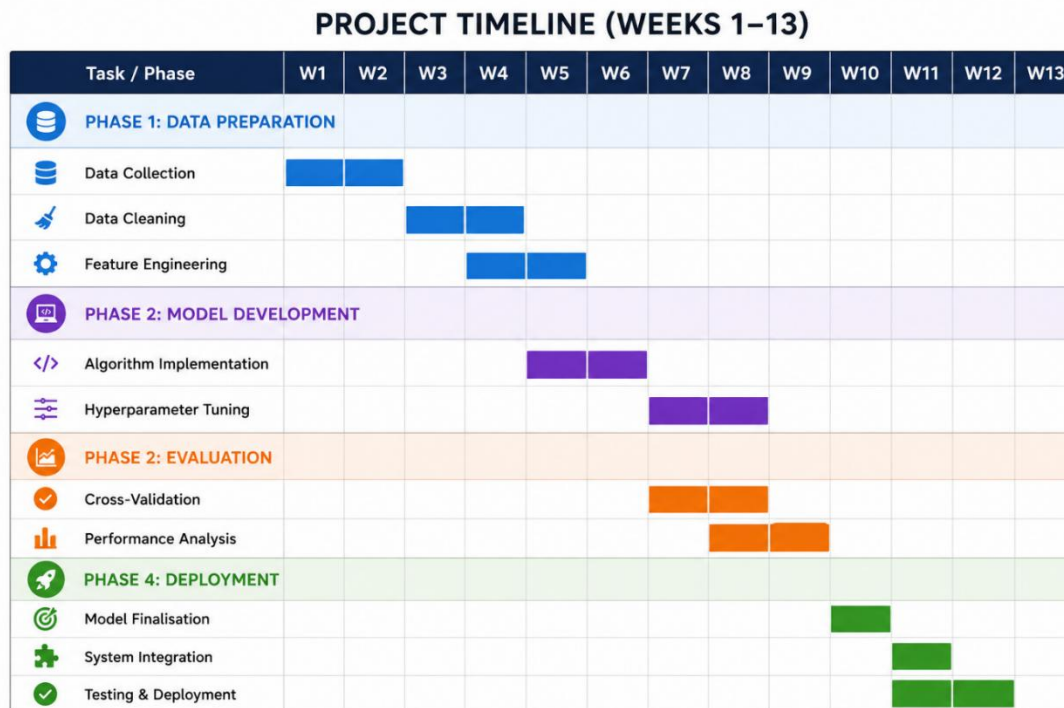


Figure 2 Gantt Chart

4. Discussion and Recommendation

4.1 Algorithm Comparison and Critical Analysis

A critical comparison and contrast of the candidate algorithms is performed using the literature review and the practical limit of the industry. Direct comparison is very important to justify the final recommendation.

4.1.1. The inadequacy of simple linear and distance-based models:

Logistic Regression: It is very interpretable and very cheap to run. But its linearity is essentially not enough to model the complex, non linear relationships between sensor readings, like how rising temperature and increasing vibration together predict failure. The research shows that its performance on complex PdM tasks is often unsatisfactory [13]. In this application, the risk of undetected incidents would be unacceptably high.

KNN is very simple theoretically and can make predictions very fast, which could be a good thing. But the “curse of dimensionality” states that as the number of features grows and, with it, the number of sensor readings too, the importance of distance between data points decreases [14]. The feature engineering phase will add more dimensions and will inconsistently affect KNN performance. Likewise, the lack of feature importance is a significant disadvantage in industrial environments where it is important to find the root cause of failure.

4.1.2. The Performance vs. Interpretability Trade-off of complex models:

Support Vector Machine: Support Vector Machines are strong classifiers that can provide high accuracy, especially for nonlinear data, by using kernel methods. But this power is not without its price. The training process is resource consuming and demands high computational power, which directly challenges the company’s limited resources. Furthermore, the kernel transformations result in a “black box” effect, which makes it very difficult for engineers to understand the model [15]. Lack of interpretability limits trust and prevents understandings from the model predictions for improved operational processes.

XGBoost often gains better predictive accuracy on a predictive maintenance scale. Regularization and the effectiveness are impressive [16]. However, in many studies, this accuracy is often insignificant compared to Random Forest, and the cost is huge.

XGBoost is a more complicated group of trees, and it is difficult to interpret without the help of additional tools such as SHAP. The added complication, longer training time, and need for an additional ‘explainer’ model all gain costs and expenses that balance the small improvement in accuracy in a resource limited situation [17]. The need for fast and transparent decisions suggests a more interpretable model is useful.

4.1.3. The Superior Balance of Random Forest:

Strong performance: Random Forest is a large set of decision trees. Its “bagging” methodology reduces overfitting and variance, thus making it significantly strong especially on noisy sensor data.

Handling Non-Linearity: Random Forest can capture complicated, non-linear decision boundaries as it builds a connection of decision trees with different subsets of features and data, without requiring data transformation. This is a direct benefit over linear models such as logistic regression and a simpler application than SVM for complex data.

Random Forest maintains consistent performance on imbalanced data. The combination voting structure can effectively address the shortage of failure events using either class weighting or SMOTE, a problem that seriously affects the accuracy of models such as logistic regression.

Fundamental Interpretability: The most attractive feature of this technique for this application is that it offers rankings of feature importance. This native “explainability” is a particular advantage over SVM and XGBoost. For the manufacturing company, recognizing that five minute vibration rise is the strongest failure indicator that delivers actionable intelligence, not just a prediction. It lets maintenance teams dig into specific issues, build trust, and learn continuously.

Comparison Summary Table

Criteria	Random Forest	XGBoost	SVM	KNN	logistic regression
Accuracy	High	Very High	Variable	Moderate	Moderate
Interpretability	Good	Fair	Poor	Good	Excellent
Training Time	Moderate	Moderate	High	Low	Very Low
Handles Nonlinearity	Excellent	Excellent	Good	Poor	Poor
Handles Imbalance	Good	Good	Fair	Good	Poor
Resource Requirements	Moderate	Moderate	High	Low	Minimal

4.2 Recommendation: Random Forest

After a comprehensive evaluation, Random Forest has been identified as the most suitable algorithm for the application. The main points of support for this proposal are as follows:

- Balanced Performance:** Random Forest is not only great performing about accuracy but also precision and recall. Besides, it can model very well the complex non-linearities present in the sensor data.
- Greater Transparency:** While XGBoost needs external instruments like SHAP for explanation, Random Forest comes with a feature importance ranking as one of its components. Given that the industrial client might be interested in determining the sensor measurements like temperature, vibration, or pressure that most strongly indicate failure, it would be much more insightful for the client to be able to draw operational insights rather than having predictions alone.
- Resilience to Imbalanced Data:** Random Forest is effective in handling imbalanced datasets that are common in failure prediction, class weighting and collection voting are the ways the method deals with the imbalanced data.

4. Computational Efficiency: Random Forest training times as short as 3 minutes on regular hardware were mentioned in the literature, and because of this, implementation will be well within the company's resource limit. It also doesn't require the heavy computation deep learning needs.

5. Operational Feasibility: Random Forest strikes a balance when it comes to the much higher performance of XGBoost and the ease of Logistic Regression. Without a heavy computational load or large data needs, it still outputs enterprise level performance.

6. Proven Track Record: Many independent research works have demonstrated the capability of Random Forest in PdM applications showing an accuracy rate of 86.4% to 100% in different cases [18].

4.3 Evaluation Recommendations

Primary Metric: The main criteria for evaluation should be the F1-score since it considers both precision and recall equally. That is more relevant in case of unbalanced datasets where the accuracy by itself can give a false impression [19]. Precision and recall should be reported separately to provide a transparent view of the tradeoff between false alarms and missed detections.

Supplementary Metric: ROC-AUC is a useful extra measure that tells one how well the model can separate the two classes as the decision threshold changes [20].

Cost Analysis: Cost sensitive evaluation using the defined cost matrix will ensure the model aligns with business objectives rather than purely statistical performance [21].

5. Conclusion

This paper showed the problem of predictive maintenance deployment in factories with limited calculation resources. The results shows the importance of selecting a suitable machine learning model to boost productivity and reduce maintenance costs. Both XGBoost and SVM gave excellent predictions but at the cost of high calculation resources. Logistic Regression and KNN were easier to use but they were not able to capture the complex patterns in the data. The best choice was Random Forest which offered a balance of accuracy interpretability and computational efficiency.

This report systematically addresses the topic of predictive maintenance in manufacturing settings with computational resource limits. The analysis stresses that choosing the right algorithm should be a planned strategic decision. For this purpose, it first reviews the existing research on the subject to perform a comprehensive assessment of five different algorithms like logistic regression, K-Nearest Neighbors, Support Vector Machine, XGBoost and Random Forest. Still, the comparison of the classifiers in the benchmarking experiment shows that XGBoost and SVM have very good predictive performance, but they require a very high level of calculation and are also black boxes, making interpreting them unreasonable. On the contrary, logistic regression and KNN are simple, yet they do not have the capacity to capture complex interaction patterns in sensor data. Out of the five algorithms considered, Random Forest wins hands down, as it is able to strike a unique balance between the three opposing demands of accuracy, interpretability, and computational efficiency. If the company implements it, a switch will be made from reactive to predictive maintenance, which will yield real advantages like less downtime, lower costs, longer equipment life, and improved operational efficiency.

Besides, it also gives the feature importance, which comes with possible involvement, as it makes the decision not only about the time of failure but also the reason behind it, thus generating confidence and driving the cycle of further enhancement.

Bibliography

- [1] A. Jadhav, "Embedded," Embedded, 19 01 2026. [Online]. Available: https://www.embedded.com/the-importance-of-predictive-maintenance-in-manufacturing/?utm_source=eetimes&utm_medium=relatedcontent. [Accessed 22 06 2026].
- [2] D. R. R. Faisal Ramzan, "EconPapers," EconPapers, 30 01 2025. [Online]. Available: https://econpapers.repec.org/article/sprsnop/v_3a6_3ay_3a2025_3ai_3a4_3ad_3a10.1007_5fs43069-025-00584-0.htm. [Accessed 18 06 2026].
- [3] Cambridge University Press, "Cambridge University Press," Cambridge University, 20 02 2026. [Online]. Available: <https://www.cambridge.org/core/journals/ai-edam/article/predictive-maintenance-in-aircraft-engine-maintenance-using-the-cmapss-dataset-performance-comparison-and-evaluation-of-machine-learning-classification-algorithms/93B83E8496618EC0DAE786CFE93909>. [Accessed 18 06 2026].
- [4] G. J. T. B. L. B. A. S. X. L. John Williams, "Science Direct," Science Direct, 01 05 2026. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0360835226001154?via%3Dihub>. [Accessed 18 06 2026].
- [5] D. N. U. Rachmat Hidayat, "Research Gate," Journal of Computer Science, 01 04 2025. [Online]. Available: https://www.researchgate.net/publication/390382728_Performance_Prediction_Model_for_Predictive_Maintenance_Based_on_K-Nearest_Neighbor_Method_for_Air_Navigation_Equipment_Facilities. [Accessed 18 06 2026].
- [6] M. K. A. G. Michał Cioch, "Advances in Science and Technology Research Journal," Lublin University of Technology, 01 10 2025. [Online]. Available: <https://www.astrj.com/pdf-208284-129233?filename=Comparison-of-machine-lea.pdf>. [Accessed 18 06 2026].
- [7] A. Balan, "Zenodo," Journal of Global Research in Multidisciplinary Studi, 17 06 2022. [Online]. Available: <https://zenodo.org/records/20735009>. [Accessed 18 06 2026].
- [8] N. R. G. V. K. S. S. D. V. S. Parthiban, "InternationalJournalssrg," SSRG International Journal of Electrical and Electronics Engineering , 14 01 2026. [Online]. Available:

<https://www.internationaljournalssrg.org/IJEEE/2026/Volume13-Issue1/IJEEE-V13I1P113.pdf>. [Accessed 18 06 2026].

- [9] R. K. Alshirawi, "Repository," Rochester Institute of Technology RIT Digital Institutional Repository RIT Digital Institutional Repository, 01 12 2025. [Online]. Available: <https://repository.rit.edu/cgi/viewcontent.cgi?article=13564&context=theses>. [Accessed 18 06 2026].
- [1 F. A. B. B. H. P. Fi Imanur Sifaunnufus Ms, "ANALYZING EEG SIGNALS FOR STRESS DETECTION
0] USING RANDOM FOREST ALGORITHM," *ANALYZING EEG SIGNALS FOR STRESS DETECTION USING RANDOM FOREST ALGORITHM*, vol. 17, pp. 8-10, 2024.
- [1 Z. U. R. I. A. & S. U. I. Uzair Hassan, "Springer Nature Link," 30 09 2025. [Online]. Available:
1] <https://link.springer.com/article/10.1007/s43926-025-00188-4>.
- [1 S. S. S. Sanagavarapu, "IEEE XPLORE," IEEE XPLORE, 19 08 2021. [Online]. Available:
2] <https://ieeexplore.ieee.org/document/9574668/citations#citations>. [Accessed 22 06 2026].
- [1 K. U. Shimpei Kawada - Presenter, "AIChE," 2019 AIChE Annual Meeting, 17 02 2019.
3] [Online]. Available: <https://proceedings.aiche.org/conferences/aiche-annual-meeting/2019/proceeding/paper/767g-extended-algorithm-generalized-linear-model-cope-multicollinearity-and-nonlinearity>. [Accessed 22 06 2026].
- [1 A. AYON, "Mendeley Data," Mendeley Data, 23 02 2026. [Online]. Available:
4] <https://data.mendeley.com/datasets/vs9jgz76yt/1>. [Accessed 22 06 2026].
- [1 T. M. a. W. Duch, "Support Feature Machines: Support Vectors are not enough.,"
5] Department of Informatics, Nicolaus Copernicus University Grudziądzka 5, 87-100 Toruń, Poland, <https://ar5iv.labs.arxiv.org/html/1901.09643>, 2026.
- [1 I. F. Merizio, F. R. Chavarette and E. F. de Almeida, "Biblioteca Virtual," ENGINEERING
6] COMPUTATIONS, 03 04 2025. [Online]. Available: <https://bv.fapesp.br/en/publicacao/278232/>. [Accessed 22 06 2026].
- [1 Y.-H. H.-H. L. S. & M. Kuo-Yi Lin, "Predictive maintenance in industrial systems: an XGBoost-
7] based approach for failure time estimation and resource optimization," *Predictive maintenance in industrial systems: an XGBoost-based approach for failure time estimation and resource optimization*, vol. 1, no. an XGBoost-based approach for failure time estimation and resource optimization, p. 899, 2025.

- [1 A. M. M. a, "National Library of Medicine," PubMed Central, 05 12 2024. [Online]. Available: 8] <https://pmc.ncbi.nlm.nih.gov/articles/PMC11667603/>. [Accessed 22 06 2026].
- [1 S. K. Yashashree Mahale, "Springer Nature Link," Springer Nature Link, 12 04 2025. [Online]. 9] Available: <https://link.springer.com/article/10.1007/s42452-025-06827-3>. [Accessed 21 06 2026].
- [2 E. R. 1. T. 1, "Science Direct," CellPress, 4 06 2024. [Online]. Available: 0] https://www.sciencedirect.com/science/article/pii/S2666389924001090?fr=RR-2&ref=pdf_download&rr=a0d2fcbafd45882e. [Accessed 22 06 2026].
- [2 S. Prasetio and G. Wibisono, "IEEE Xplore," IEEE, 15 03 2024. [Online]. Available: 1] <https://ieeexplore.ieee.org/document/11236443>. [Accessed 22 06 2026].
- [2 L. Zindel, "Wiss," Wiss, 11 03 2026. [Online]. Available: [https://wiss.com/predictive-2\] maintenance-roi-cost-savings-for-manufacturers/](https://wiss.com/predictive-maintenance-roi-cost-savings-for-manufacturers/). [Accessed 18 06 2026].