

Article

Predictive Maintenance in Maritime Operations: A Data-Driven Approach for Early Warning Systems

Nick Z. Zacharis ^{1,*}, Kyriakos N. Sgarbas ², George Leventakis ³, Petros Savvidis ³, Dimitris Papachristos ³, Nikitas Nikitakos ⁴ and Iosif Progoulakis ⁴

¹ Department of Informatics and Computer Engineering, University of West Attica, Ag. Spyridonos Str., Egaleo, 12243 Athens, Greece

² Department of Electrical and Computer Engineering, University of Patras, 26504 Rio, Greece; sgarbas@upatras.gr

³ SKYLD Security and Defense Ltd., Alkaiou 23, Egkomi, 2404 Nicosia, Cyprus; euprojects@skyld.com.cy (G.L.); p.savvidis@skyld.com.cy (P.S.); dimpapachristos@yahoo.gr (D.P.)

⁴ Department of Shipping Trade and Transport, University of the Aegean, 2a Korai Street, 82132 Chios, Greece; nnik@aegean.gr (N.N.); i.progoulakis@aegean.gr (I.P.)

* Correspondence: nzach@uniwa.gr; Tel.: +30-210-538-5394

Abstract

This study outlines a comprehensive analytical methodology applied to a vessel's operational data to establish a rigorously validated, data-driven baseline for the Marine and Health Usage Monitoring System (MHUMS) algorithmic modules. By correlating a continuous three-month dataset of sensor telemetry with historical alarm logs, the primary objective was to develop and train a robust predictive model. This study addresses the critical challenge of converting continuous marine sensor telemetry into actionable early warnings by establishing a validated predictive baseline. The proposed baseline predictive architecture achieved a Recall of 0.924 and an area under the ROC curve (AUC) of 0.939 for a 15 min forward prediction window, proving the feasibility of transitioning from reactive diagnostic monitoring to proactive early warnings.

Keywords: predictive maintenance; shipping 4.0; machine learning; MHUMS; maritime digitalization; sensor telemetry; early warning systems

1. Introduction

The global maritime industry is undergoing a digital transformation known as Shipping 4.0, characterized by the integration of the Internet of Things (IoT), Big Data analytics, and Artificial Intelligence (AI). Traditional maintenance strategies, often reactive or calendar-based, are increasingly inefficient, leading to unexpectedly high Operational Expenditure (OPEX), where maintenance and repair activities alone can account for up to 40% of the total operating costs and unexpected downtime [1].

The modern maritime industry increasingly relies on advanced data analytics to monitor vessel health, reduce downtime, and prevent critical equipment failure [2,3]. Traditional planned maintenance is rapidly being superseded by Condition-Based Monitoring (CBM) and Predictive Maintenance (PdM) paradigms, driven by the proliferation of Internet of Things (IoT) sensors onboard modern vessels [4,5]. However, despite these technological advancements, bridging the gap between raw, continuous sensor telemetry and actionable, early-warning predictions in real-world, noisy marine environments remains a significant challenge due to specific maritime data acquisition constraints: (i) temporal



Academic Editor: Yordan Garbatov

Received: 7 May 2026

Revised: 20 June 2026

Accepted: 25 June 2026

Published: 6 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

misalignment between per second discrete alarm logs and per-minute continuous sensor streams, (ii) severe data fragmentation and packet loss caused by intermittent satellite communications at sea, and (iii) high ambient and operational noise generated by large two-stroke marine engines, which masks early degradation signatures.

The MHUMS project addresses the aforementioned challenges. The system's primary goal is to enable Condition-Based Maintenance (CBM) by providing real-time monitoring and predictive insights into critical vessel assets such as propulsion, power generation, and fuel systems. Moreover, one of the main objectives of the MHUMS algorithmic development is to establish a data-driven baseline by integrating continuous sensor measurements with historical alarm logs from real-world pilot vessel data. The scope of this implementation encompasses the following:

- *Data Acquisition and Preprocessing:* In order to handle the severe heterogeneity and noise found in real-world marine data streams [6], this study implements discrete alarm downsampling to a per-minute indication, chronological timestamp alignment, and linear interpolation for telemetry gaps, to ensure high-fidelity input.
- *Feature Extraction and Anomaly Detection:* Instead of relying on manual, expert-defined thresholds that ignore underlying physical redundancies and cross-sensor correlations, this work utilizes hierarchical clustering on sensor features to automatically isolate coupled subsystems and identify structural patterns and system redundancies to inform feature engineering [7].
- *Predictive Maintenance Inference:* Most current commercial systems operate purely as reactive diagnostic tools rather than early-warning mechanisms, or they skip straight to highly complex Remaining Useful Life (RUL) models [8] without establishing a reliable baseline. This work explicitly builds an early-warning classification baseline optimized for extreme class imbalance.

The core problem statement addressed by this work is the lack of a reliable, mathematically transparent, and structurally validated baseline that can ingest heterogeneous, asynchronous per second alarm logs and continuous per-minute sensor telemetry to issue actionable early warnings 15 min before critical equipment failure. To address this problem, the authors deploy a unified data engineering pipeline within the MHUMS framework that synchronizes temporal resolutions, optimizes high-dimensional input spaces via hierarchical sensor clustering, and evaluates regularized classifiers optimized for heavy class imbalance.

2. Materials and Methods

This section presents the dataset, the MHUMS system architecture, and the methodology, i.e., data acquisition, preprocessing, feature engineering, and the implementation of the machine learning predictive model.

2.1. The Vessel and MHUMS System Architecture

The Marine and Health Usage Monitoring System (MHUMS) employs a hybrid Edge-Cloud architecture designed to monitor critical vessel assets, including propulsion, power generation, and fuel systems. For the purposes of this study, a continuous three-month dataset of operational telemetry was collected from a real-world pilot vessel, i.e., the *Orient Pearl* (IMO 9450727, MMSI 636014610), a bulk carrier under the management of *Interorient Shipmanagement* (142 Franklin Roosevelt, Limassol, Cyprus 3011). Launched in 2010, the vessel has a length of 180 m, a beam of 30 m, a gross tonnage (GT) of 23,456, and a deadweight tonnage (DWT) of 34,403, designed specifically for international dry bulk cargo transport. This ship is currently the central platform for the MHUMS prototype pilot study, which utilizes an IoT-based Health and Usage Monitoring System to improve fuel efficiency,

operational diagnostics, and predictive maintenance capabilities. The framework integrates an array of onboard sensors to monitor engine performance, fuel usage, lubrication integrity, and environmental variables. Ultimately, this sensor data feeds into AI-driven analytics to provide the crew with real-time alerts and intelligent decision support.

The MHUMS architecture follows a hybrid Edge-Cloud model to balance real-time response with fleet-wide analytics. The system is divided into seven core functional modules that manage the data lifecycle:

- *F1 and F2 (Data Acquisition and Data Manipulation)*: Interfaces with Voyage Data Recorders (VDRs) and Automation and Monitoring Systems (AMSs) to collect raw data, applying noise filtering and time-synchronization.
- *F3 and F4 (State Detection and Health Assessment)*: Identifies the vessel's operational state and evaluates the condition of critical components.
- *F5 (Prognostics Assessment)*: Forecasts future health states and RUL using historical trends and fuel consumption metrics;
- *F6 (Advisory Generation)*: Produces actionable warnings and maintenance suggestions for the crew.
- *F7 (Virtual Assistant)*: Facilitates human-machine interaction through natural language processing.

Onboard hardware includes Local Signal Acquisition Units (SAUs) that process both analog (4–20 mA current loops) and digital (Modbus, CAN bus, NMEA) signals. An on-board Edge Server executes AI models locally, ensuring functionality even in disconnected maritime environments.

Figures 1 and 2 show the pillars-objectives-modules hierarchy and the system architecture of the MHUMS, respectively.

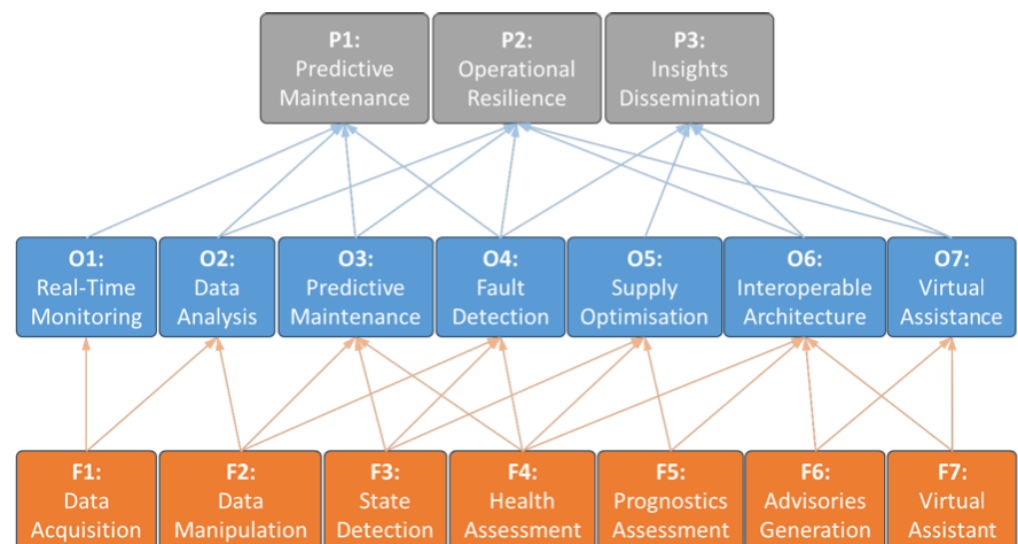


Figure 1. Pillars-objectives-modules hierarchy of the MHUMS.

2.2. Data Acquisition, Preprocessing, and Feature Engineering

Creating a reliable predictive model requires highly structured, harmonized data. As noted by Kirketerp-Møller et al. [9], MLOps pipelines in maritime settings depend heavily on data integrity. The pilot analysis utilized an operational dataset collected over three months (September to December 2025), comprising four primary data sources:

- Threshold metadata for 277 alarms.
- A structured log of 70,808 alarm activation instances.

- A sensor catalog detailing 186 sensors.
- The continuous time-series baseline from 157 sensor tags across 130,494 time instances.

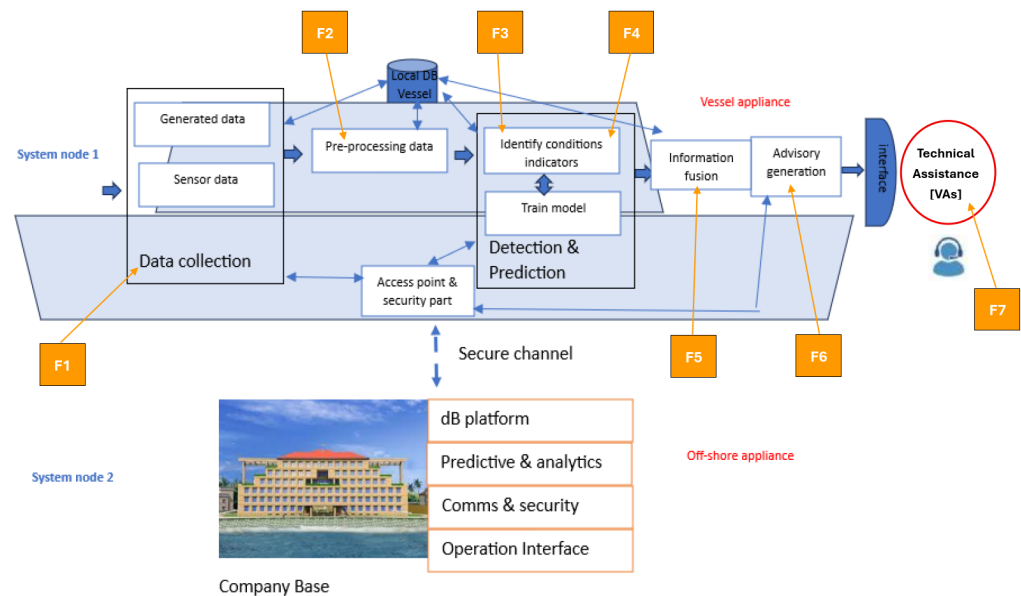


Figure 2. The MHUMS architecture, pinpointing the functional modules (F1–F7) on the ship.

A significant engineering challenge involved aligning heterogeneous logs. Alarm events were recorded with per second precision, while sensor time-series data possessed per-minute precision causing temporal synchronization issues frequently observed in maritime IoT streams [4]. To synchronize the distinct temporal resolutions, discrete alarm logs (recorded with per second precision) were downsampled to a per-minute binary state. If an alarm instance was active at any point within a given 60 s window, that minute was flagged as an active ‘Alert’ state; otherwise, it was classified as ‘Normal’. Continuous sensor time-series data (recorded at per-minute intervals) were then directly joined to these synchronized windows using standardized timestamps.

The dataset contained raw sensor per-minute measurements paired with historical alarm logs provided by the vessel’s control systems. Given the proprietary nature of the maritime operations, the data collection adhered to strict confidentiality protocols.

Since maritime environments inherently produce noisy and fragmented data, to prepare the dataset for machine learning ingestion, the Orange Data Mining Toolbox [10] was used for data preprocessing. The following preprocessing steps were implemented:

- *Time Synchronization:* A critical step involved aligning the continuous sensor readings with the discrete timestamps of the historical alarm logs to create labeled instances for supervised learning.
- *Data Cleaning:* Missing values and sensor transmission errors, common in maritime telemetry, were addressed using standard interpolation techniques.
- *Feature Extraction:* Structural patterns and system redundancies were analyzed to extract meaningful features. By evaluating the correlation between specific sensor fluctuations and subsequent alarms, the input dimensionality was optimized, focusing on the most critical early-warning indicators.

2.3. Predictive Modeling and Algorithm Selection

The transition from diagnostic to predictive maintenance requires algorithms capable of handling non-linear, high-dimensional data. For this baseline model, the Orange Data Mining Toolbox [10] was used again. Four machine learning algorithms were used and compared: namely, Tree, Random Forest, Logistic Regression, and Neural Network. The algorithms produced trained models to classify operational states into “Normal” or “Alert” based on the extracted features. The analytical pipeline, including model training, cross-validation, and performance evaluation, was implemented as an Orange dataflow (Figure 3). The best performing model was selected as a proof of concept validating the methodology.

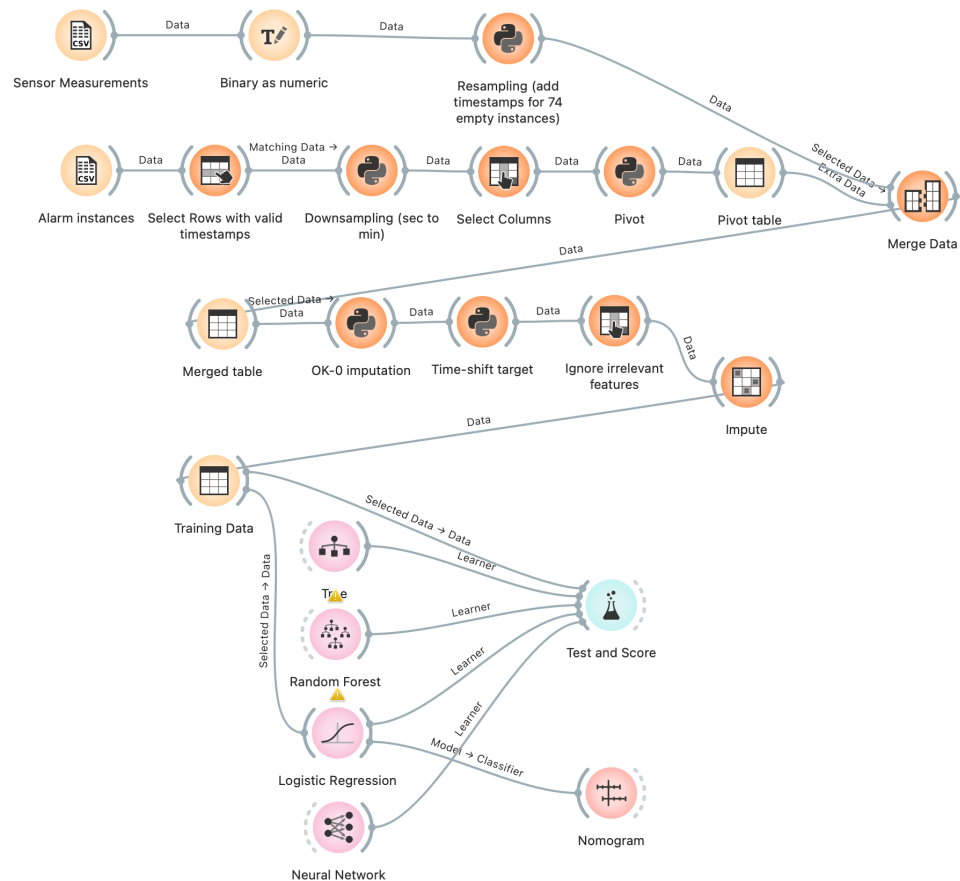


Figure 3. The Orange data mining platform workflow used for the preprocessing and training process.

2.4. Methodological Declarations

- **Data Availability:** The raw telemetry and alarm datasets analyzed in this study are derived from a commercial pilot vessel. Due to confidentiality agreements and operational security policies, the raw datasets cannot be made publicly available. However, the methodology described and the proof of concept presented in this report may allow replication using similar maritime datasets.
- **Ethical Approval:** This study involves the computational analysis of machine-generated sensor data. No human or animal subjects were involved; hence, ethical approval was not required.
- **Generative AI Disclosure:** During the preparation of this manuscript, a Generative AI tool (Gemini 3) was utilized for text editing to improve readability and grammar. The authors have thoroughly examined the final output to ensure that no hallucinations or other type of errors were introduced.

3. Results

This section details the outcomes of the data analysis and the performance of the predictive model. The evaluation focuses on the model's ability to accurately classify operational states and provide reliable early warnings for impending critical alarms.

3.1. Exploratory Data Analysis and Feature Structuring

To evaluate the information content of the massive telemetry matrix, Principal Component Analysis (PCA) [11] was utilized to handle the high dimensionality inherent to marine engine telemetry [12]. Missing values were imputed, and strict data standardization ($\mu = 0$, $\sigma^2 = 1$) was applied. As shown in Figure 4, the PCA revealed high redundancy in the sensor space, indicating that more than 80% of the dataset's variance could be captured using only 10 principal components.

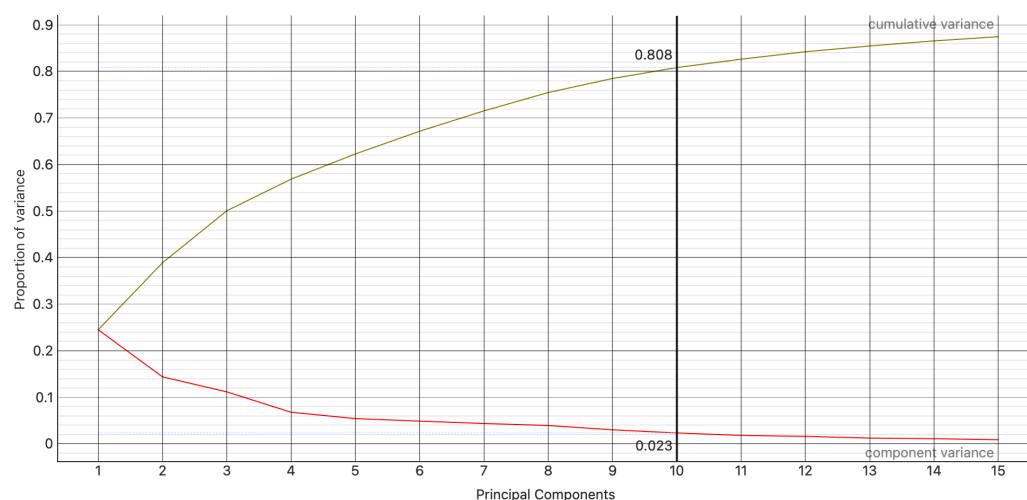


Figure 4. PCA revealed that more than 80% of the data variance can be captured using only 10 principal components.

Following PCA, Louvain clustering grouped the time-instances into 331 clusters. Louvain clustering [13] is a heuristic network-based algorithm designed to optimize graph modularity to detect community structures. In this analysis, it was applied to a k-nearest neighbors (k-NN) graph constructed from the time-instances to uncover hidden global operational modes. However, due to high temporal autocorrelation, the resulting 331 clusters merely segmented chronological timelines rather than distinct systemic states, prompting a pivot to column-based hierarchical clustering. Consequently, the analytical focus pivoted to mapping physical redundancy by performing Hierarchical Clustering directly on the sensor columns. Using Pearson correlation distance combined with complete linkage, the algorithm successfully isolated physically coupled subsystems, an essential step for explainable anomaly detection [14]. A height ratio of 17% on the resulting dendrogram of the Hierarchical Clustering yielded 49 clusters, as shown in Figure A1.

3.2. Methodological Approach to Predictive Modeling

The core objective of MHUMS is generating actionable early warnings. Frequency analysis revealed severe sparsity in alarm activations; only alarms with over 100 activation instances were deemed acceptable for modeling to mitigate class imbalance. Of these, the alarm with tag MA14 (M/E NO.4 SCAV.AIR FIRE DET.), an analog measurement ideal for observing continuous degradation in two-stroke engines, was selected as the primary test case [9,12]. The activation pattern of MA14 is shown in Figure 5. The predictive model aimed to foresee the activation of MA14 fifteen minutes before its occurrence, emphasizing recall.

Therefore, the dataset was structurally transformed to enable machine learning ingestion:

- *Feature Transformation:* The long-format alarm log was pivoted into a wide-format matrix using standardized timestamps.
- *Activation Level Normalization:* The MA14 alarm actually has two activation levels: H (high) and HH (high-high). The model was trained to handle both equally.
- *Safe Imputation:* Null values were safely imputed with zeros, translating the absence of an alarm into a definitive “Normal” class. Records with invalid timestamps were omitted.
- *Target Shifting:* To transition from diagnostic to predictive modeling, a 15 min forward time shift was applied to the MA14 threshold, consolidating multiple severe states into a binary “Alert” category.

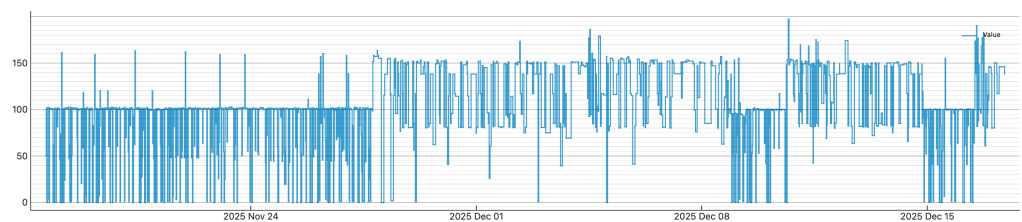


Figure 5. Activation of the MA14 alarm. The horizontal axis shows timestamps; the vertical axis shows temperature.

Using the aforementioned dataset, several supervised baselines were trained, utilizing specific hyperparameters to mitigate class imbalance, reflecting methodologies validated in the recent literature [3,5]. These included Decision Trees, Random Forests, Logistic Regression with Ridge (L2) regularization, and Artificial Neural Networks (MLPs) designed to map complex, non-linear sensor interactions [15]. Specifically,

- *Decision Tree:* Configured as a highly interpretable baseline model. To prevent the tree from overfitting and simply memorizing random sensor noise, strict pre-pruning parameters were applied: a maximum depth of 10 levels, a requirement of at least 15 instances per leaf, and a rule to halt splitting if a node reached a 95% majority class consensus. Subsets smaller than 50 instances were restricted from further splitting to ensure statistical validity.
- *Random Forest:* An ensemble approach utilized to improve generalization. The model was constructed using 100 independent decision trees. Crucially, a ‘balanced’ class distribution parameter was enforced, which automatically adjusts the algorithmic weights inversely proportional to class frequencies, forcing the trees to pay equal attention to the rare ‘Alert’ class.
- *Logistic Regression:* Selected to capture linear degradation trends in the analog sensors. A Ridge (L2) regularization penalty with a baseline inverse regularization strength ($C = 1$) was applied. This is particularly vital for ship telemetry, as the L2 penalty stabilizes the model against the high multicollinearity (heavily correlated sensors) present in the data. Like the Random Forest, class weights were explicitly balanced.
- *Artificial Neural Network (Multi-Layer Perceptron):* Included to capture highly complex, non-linear sensor interactions. The architecture consisted of a single hidden layer with 100 neurons, utilizing the Rectified Linear Unit (ReLU) activation function to prevent gradient vanishing, and the Adam solver for efficient stochastic optimization. Training was capped at 200 iterations to prevent over-convergence on the training set.

Both the preprocessing and the training stages were implemented using the Orange data mining platform [10]. Figure 3 presents the actual dataflow.

3.3. Model Evaluation and Operational Metrics

In the same dataflow (Figure 3), the Test and Score widget was used to evaluate the performance of each model. The results are shown in Table 1. The table depicts six metrics (AUC, CA, F1, Prec, Recall, MCC) for each model (algorithm). Their details and their significance are explained below:

- *CA (Classification Accuracy):*

$$CA = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

The percentage of total predictions (both “Alert” and “Normal”) that the model got correct. In our dataset, alarms are incredibly rare. If 99% of the data is “Normal,” a broken model could simply guess “Normal” every single time, missing every single real alarm, and still achieve a 99% CA. Therefore, for this specific project, CA is not a reliable metric due to class imbalance.

- *Recall (also known as sensitivity):*

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Out of all the actual alarms that occurred, what percentage did the model successfully find? If the model has a Recall of 0.90, it means it successfully warned about 90% of the impending MA14 fires/faults, but missed 10% (as False Negatives). In predictive maintenance, Recall is usually the most important metric because missing a critical engine alarm is far more expensive and dangerous than dealing with a false alarm.

- *Prec (Precision):*

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Out of all the times the model predicted an alarm would happen, how many times was it actually right? If the model has a Precision of 0.50, it means that half the time it sounds the “Alert” bell, nothing is actually wrong (these are False Positives), while missing a fire (low Recall) is bad, having too many false alarms (low Precision) leads to “alarm fatigue,” where the crew starts ignoring the system.

- *F1 (F1-Score):*

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

The harmonic mean of Precision and Recall. It combines both metrics into a single number. Because it is very easy to build a model with perfect Recall (just predict “Alert” every single minute) or perfect Precision (only predict “Alert” when it is 100% certain), the F1 score forces a balance. A high F1 score proves that the model is effectively catching the alarms (Recall) without driving the crew crazy with false warnings (Precision).

- *AUC (area under the ROC curve):* AUC measures the model’s overall ability to distinguish between the “Alert” class and the “Normal” class across all possible probability thresholds. It is calculated by integrating the area under the ROC curve. A score of 0.5 means the model is just guessing randomly. A score of 1.0 means it distinguishes them perfectly. This is a great “big picture” metric. A high AUC (e.g., >0.85) proves that the underlying data contain a strong, valid signal separating normal operations from alarm precursors, regardless of how strict or lenient we set the final alarm threshold.

- *MCC (Matthews Correlation Coefficient):*

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

A highly robust statistical rate that evaluates all four outcomes (True Positives, True Negatives, False Positives, and False Negatives). It ranges from -1 (total disagreement) to 0 (random guessing) to $+1$ (perfect prediction). In case of massive class imbalance (130,000 normal minutes vs. maybe 100 alarm minutes), F1 and AUC can sometimes still be slightly overly optimistic. MCC is mathematically known as the absolute strictest, most honest metric for severely imbalanced data. A high positive MCC score is the ultimate proof that the model genuinely understands the alarm patterns.

Table 1. Results of the evaluation procedure.

MODEL	AUC	CA	F1	Prec	Recall	MCC
Tree	0.934	0.944	0.039	0.341	0.020	0.072
Random Forest	0.926	0.937	0.138	0.242	0.092	0.131
Logistic Regression	0.939	0.886	0.470	0.315	0.924	0.501
Neural Network	0.939	0.944	0.025	0.297	0.013	0.053

Given the extreme rarity of fault instances in maritime telemetry, standard Classification Accuracy is fundamentally unreliable; a model guessing “Normal” constantly could achieve 99% accuracy while missing critical failures [7]. Therefore, model evaluation prioritized Recall in order to ensure minimal missed faults (false negatives) and the F1-Score and MCC, to guarantee a balanced precision that minimizes false positive alarm fatigue.

As shown in Table 1, Logistic Regression outperformed the other models due to its Recall (also having the best F1 and MCC scores). The nomogram of Figure 6 highlights the sensors that contributed most to the Logistic Regression prediction.

For explainability, feature contributions were extracted. Sensors such as ME1 Density and ME1 THRUST BEARING TEMP emerged as important contributing factors driving the predictive logic. The need for explainability is paramount in maritime AI to foster operator trust, a factor heavily emphasized by Kalafatelis et al. [16], and the nomogram of a logistic regression model is a quite convenient self-explainable representation.

The information on the nomogram is interpreted as follows:

- *Y-Axis Labels:* These represent individual, raw physical sensor tags in hexadecimal notation (e.g., 00010106, 02FF0305) rather than reduced PCA components. They are ranked top-to-bottom by their relative importance.
- *Top Axis (“Points”):* This is a standardized scale ranging from 0 to 100. It defines the universal scoring contribution that any given feature value imparts to the predictive model.
- *Cartesian Space Lines (horizontal feature axes):* These display the actual physical measurement ranges of each sensor. Their scales and ranges differ completely because they reflect distinct physical units (e.g., temperatures in $^{\circ}\text{C}$ vs. density in kg/m^3). To read the model, an operator locates a sensor’s current value on its specific line and projects a vertical line upward to the top axis to find its corresponding “Points” score.
- *Bottom Axis (“Total Points” and “Probabilities”):* The “Total Points” line accumulates the scores from all individual sensors. This sum is mapped directly to the final “Probabilities (%)” scale beneath it, representing the calculated risk of the MA14 alarm activating within the next 15 min.

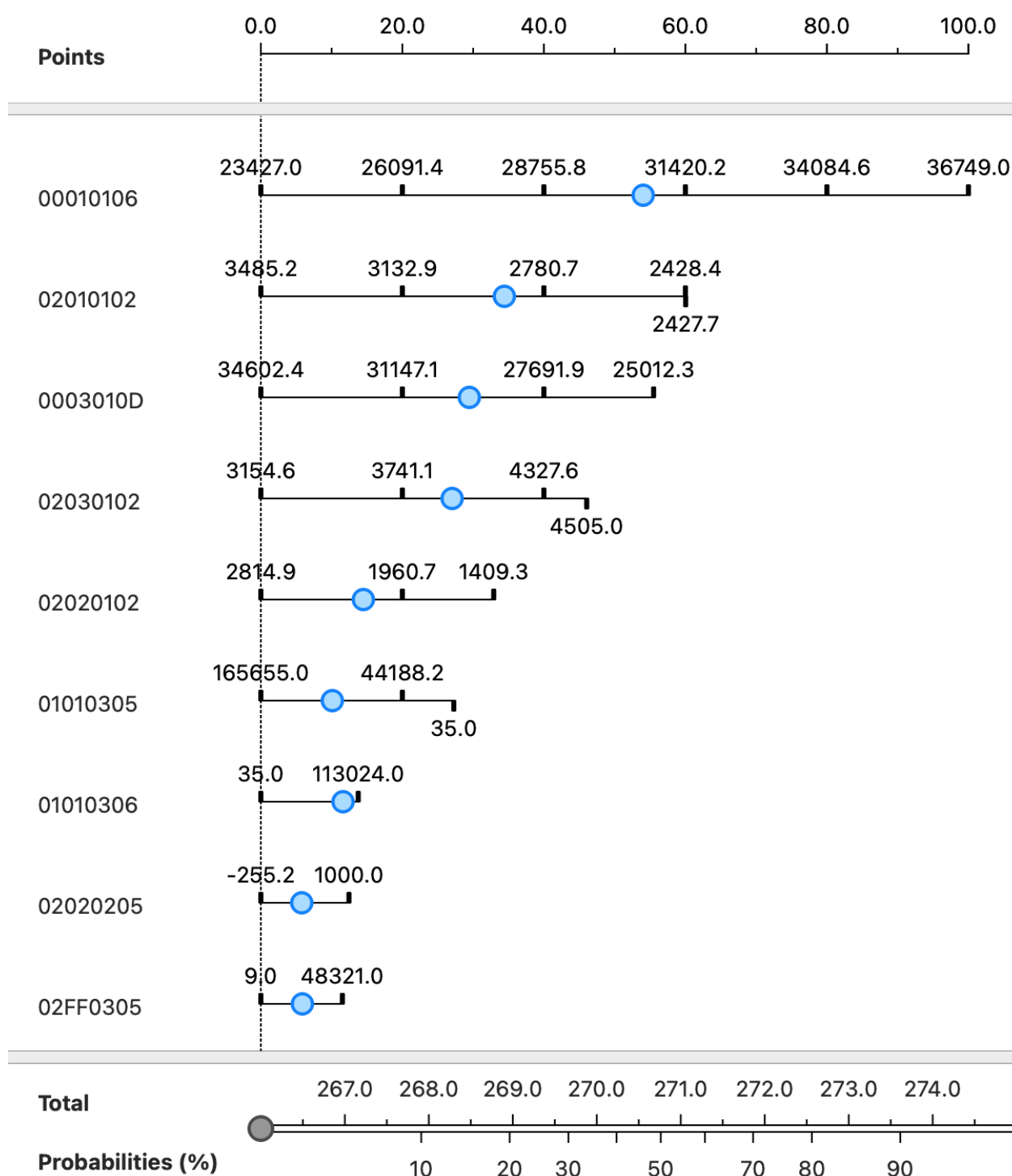


Figure 6. A nomogram depicting the sensors that contributed most to the Logistic Regression prediction (output of the Orange platform).

3.4. Algorithmic Integration and Deployment Architecture

To ensure zero-latency processing without relying on intermittent satellite communications, the predictive model can be deployed directly onto the onboard Edge Server of the MHUMS. Operating as a containerized service, it continuously ingests real-time sensor streams and broadcasts predictions via APIs, a deployment strategy highly relevant for both active vessels and shipyards utilizing AI-driven oversight [17].

Outputs are packaged into standardized JSON payloads, enforcing a strict Data Contract with downstream interfaces. Each payload mandates the timestamp, the forecasted probability score for the subsystem, and a ranked vector of contributing sensors, explicitly telling the UI why an alert is triggering to enable rapid root-cause analysis.

However, since this is still a pilot (proof-of-concept) application, the process will be repeated via telemetry to accumulate more training data and to integrate more alarms before it is deployed.

4. Discussion

The results of this study validate the hypothesis that a data-driven approach, leveraging machine learning on continuous sensor telemetry, can establish a reliable baseline for early-warning systems in maritime operations. The transition from reactive to predictive maintenance is not merely a technical upgrade but a strategic necessity within the Shipping 4.0 framework.

4.1. Interpretation of Findings and Model Performance

The high recall and F1-score achieved by the Logistic Regression model suggest that the correlation between raw sensor fluctuations and historical alarm logs contains identifiable signatures of system degradation. Unlike traditional diagnostic systems that trigger alarms only after a threshold is breached, the MHUMS baseline model identifies these patterns in the pre-alarm phase.

One of the key strengths of the Logistic Regression model in this context is its simplicity and its robustness against the inherent noise of marine telemetry. As noted in the methodology, maritime data are often fragmented due to transmission errors or environmental interference.

However, although the experiment was limited to a single alarm, as a proof of concept, the relatively simple nature of the model allowed for a stable classification performance, confirming that feature importance analysis can effectively prioritize sensors critical for each particular alarm.

4.2. Comparison with Previous Studies

Our findings align with recent research emphasizing the potential of AI-driven predictive maintenance in the maritime sector. For instance, while other studies like refs. [2,16] explore more complex predictive architectures, our study demonstrates that simple, well-tuned models could provide robust and more computationally efficient baselines for initial deployment, if the prediction window is modest.

Furthermore, the integration of real-world pilot vessel data addresses the need for practical validation of AI models in operational environments, while other studies rely on simulated datasets, the MHUMS framework utilizes high-fidelity telemetry, proving that predictive insights can be extracted despite the complexity of real-world maritime “noisy” data.

4.3. Practical Implications and Future Directions

The implications of this research extend to the broader context of maritime digitalization. By reducing unexpected downtime and optimizing maintenance schedules, the proposed system directly addresses the 40% operational expenditure (OPEX) challenges highlighted in the introduction.

However, several areas for future research remain:

- *Scalability*: The next version of the MHUMS prototype will employ the prediction of many different alarms at various time-windows. This requires training on even more data.
- *Specialization–Generalization*: Predictive models can be specialized on a certain vessel or generalized for many vessels of a fleet. A next version of the MHUMS prototype will be trained in fleet data and it will be tested in terms of accuracy and recall against all vessels of the fleet.
- *Edge-Cloud Integration*: Future iterations of the MHUMS ecosystem will focus on deploying these models at the Edge. This will allow for real-time inference onboard the vessel, reducing the latency associated with cloud-based processing and ensuring continuous monitoring even during satellite connectivity gaps.
- *Remaining Useful Life (RUL)*: While this study focuses on early-warning classification, the next phase of research will aim to transition from fault detection to RUL forecasting, providing precise estimates of when a component will fail.

In summary, this baseline model serves as the foundational layer for the MHUMS predictive suite, demonstrating that data-driven early warnings are both feasible and accurate in modern marine engineering.

5. Conclusions

This study successfully established a rigorously validated, data-driven predictive baseline for the Marine and Health Usage Monitoring System (MHUMS) by applying machine learning techniques to real-world maritime sensor telemetry. This study demonstrated that transitioning from traditional, reactive diagnostic monitoring to an AI-driven [18], proactive early-warning system is both technically feasible and operationally effective.

Quantitatively, the top-performing regularized Logistic Regression framework proved the technical and operational feasibility of AI-driven early warnings, achieving an area under the ROC curve (AUC) of 0.939, a critical fault Recall of 0.924, and a highly strict Matthews Correlation Coefficient (MCC) of 0.501 under extreme class imbalance conditions (130,494 time instances versus sparse target activations).

While this proof of concept demonstrates that pre-alarm degradation signatures can be extracted from noisy telemetry, certain limitations remain: this study was restricted to a single continuous alarm type evaluated on a single pilot vessel over a three-month horizon, relying primarily on logistic baseline classification modeling.

Future work will directly address these constraints by: (1) scaling the training pipeline across heterogeneous multi-vessel fleet-wide datasets to evaluate model generalization; (2) implementing more complex, non-linear deep temporal sequence networks (such as hybrid BiLSTM-Attention configurations) to extend prediction horizons beyond 15 min; (3) executing direct containerized field tests of the model on onboard Edge hardware; and (4) transitioning from binary anomaly flags to definitive multi-step Remaining Useful Life (RUL) prognostications.

Author Contributions: Conceptualization, N.Z.Z. and K.N.S.; methodology, N.Z.Z. and K.N.S.; software, N.Z.Z. and K.N.S.; validation, N.Z.Z. and K.N.S.; investigation, G.L., P.S. and D.P.; resources, G.L., P.S., D.P., N.N. and I.P.; data curation, G.L., P.S., D.P. and N.N.; writing—original draft preparation, N.Z.Z. and K.N.S.; writing—review and editing, G.L., P.S., D.P., N.N. and I.P.; supervision, G.L. and I.P.; project administration, G.L. and I.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation Programmes for Research, Technological Development and Innovation “RESTART 2016–2020”, Pillar I. Smart Growth, Programme

CO-DEVELOP, RIF Proposal Number CODEVELOP-AG-SH-HE/0823/0124, MHUMS—Maritime Health Usage Monitoring System. The consortium consisted of SKYLD Ltd., University of Nicosia Research Foundation (UNRF), University of West Attica (UNIWA), and Digital Ocean Technologies (DOT) Ltd.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request. Access is restricted due to the sensitive nature of the research, the operational security considerations related to the assets involved, and the funding conditions imposed by the European Regional Development Fund and the Republic of Cyprus.

Acknowledgments: During the preparation of this manuscript, the authors used Gemini 3 for the purposes of typesetting and proofreading the manuscript. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: Authors George Leventakis, Petros Savvidis and Dimitris Papachristos were employed by the company SKYLD Security and Defense Ltd., Nicosia, Cyprus. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AMS	Automation and Monitoring Systems
AUC	Area Under Curve
CA	Classification Accuracy
CBM	Condition-Based Monitoring
DWT	Deadweight Tonnage
FN	False Negative(s)
FP	False Positive(s)
GT	Gross Tonnage
IoT	Internet of Things
JSON	JavaScript Object Notation
MCC	Matthews Correlation Coefficient
MHUMS	Marine and Health Usage Monitoring System
MLP	Multi-Layer Perceptron
OPEX	Operational Expenditure
PCA	Principal Component Analysis
PdM	Predictive Maintenance
ReLU	Rectified Linear Unit
RUL	Remaining Useful Life
SAU	Signal Acquisition Units
TN	True Negative(s)
TP	True Positive(s)
UI	User Interface
VDR	Voyage Data Recorders

Appendix A

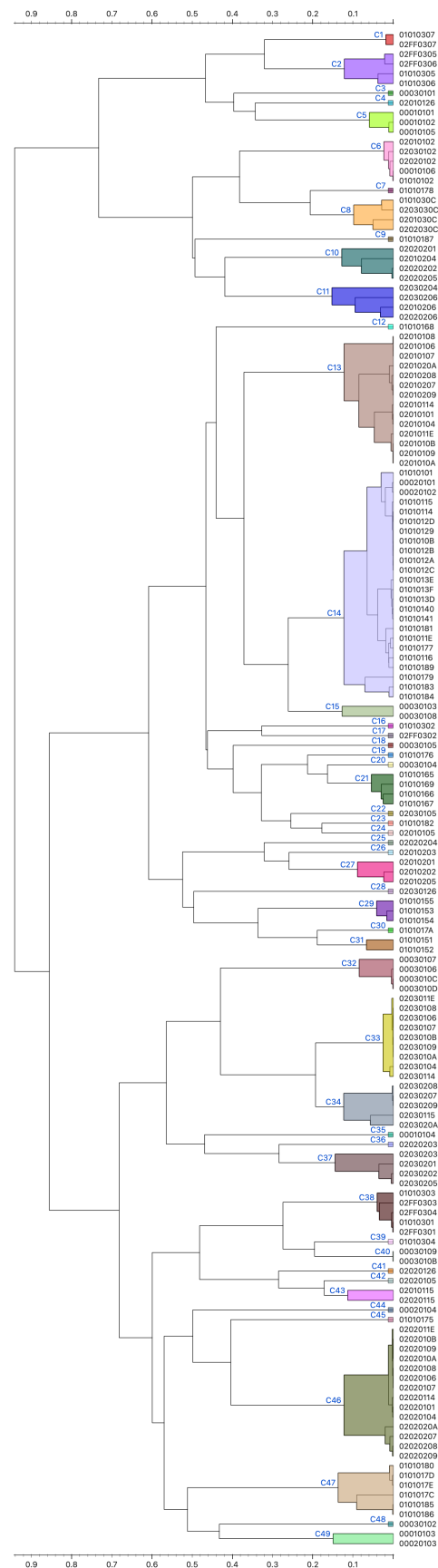


Figure A1. Hierarchical clustering with Pearson correlation distance yielding 49 clusters.

References

1. Stopford, M. *Maritime Economics*, 3rd ed.; Routledge: London, UK, 2009.
2. Kalafatelis, A.S.; Nomikos, N.; Giannopoulos, A.; Trakadas, P. A Survey on Predictive Maintenance in the Maritime Industry Using Machine and Federated Learning. *TechRxiv* **2024**. [[CrossRef](#)] [[PubMed](#)]
3. Kalafatelis, A.S.; Nomikos, N.; Giannopoulos, A.; Alexandridis, G.; Karditsa, A.; Trakadas, P. Towards Predictive Maintenance in the Maritime Industry: A Component-Based Overview. *J. Mar. Sci. Eng.* **2025**, *13*, 425. [[CrossRef](#)]
4. Setiawan, A.; Handoko, W.; Onn, C.W.; Widyaningsih, U. IoT-Based Machine Learning for Predictive Maintenance of Ship Engines: Efficiency Gains in Maritime Digitalization. *Din. Bahari* **2025**, *6*, 83–99. [[CrossRef](#)]
5. Tsallis, C.; Papageorgas, P.; Piromalis, D.; Munteanu, R.A. Application-Wise Review of Machine Learning-Based Predictive Maintenance: Trends, Challenges, and Future Directions. *Appl. Sci.* **2025**, *15*, 4898. [[CrossRef](#)]
6. Faggioni, N.; Caviglia, A.; Guarnera, N.; Schininà, E.; Sansebastiano, E.; Chiti, R. Enhancing Predictive Maintenance in the Maritime Industry with Unsupervised Learning. In *Proceedings of the International Ship Control Systems Symposium (iSCSS)*; Institute of Marine Engineering, Science and Technology (IMarEST): London, UK, 2024. [[CrossRef](#)] [[PubMed](#)]
7. Kim, D.; Lee, S.; Lee, J. An Ensemble-Based Approach to Anomaly Detection in Marine Engine Sensor Streams for Efficient Condition Monitoring and Analysis. *Sensors* **2020**, *20*, 7285. [[CrossRef](#)] [[PubMed](#)]
8. Makridis, G.; Kyriazis, D.; Plitsos, S. Predictive maintenance leveraging machine learning for time-series forecasting in the maritime industry. In *Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*; IEEE Press: Piscataway, NJ, USA, 2020; pp. 1–8. [[CrossRef](#)]
9. Kirketerp-Møller, T.; Hyldgaard, M.W.; Cai, J.; Dodis, A.I.; Rytter, N.G.M. Data-driven predictive maintenance for two-stroke marine diesel engines using machine learning and MLOps. *J. Ocean. Eng. Sci.* **2026**, *11*, 278–296. [[CrossRef](#)]
10. Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočevár, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A.; et al. Orange: Data Mining Toolbox in Python. *J. Mach. Learn. Res.* **2013**, *14*, 2349–2353.
11. Jolliffe, I.T. *Principal Component Analysis*, 2nd ed.; Springer: New York, NY, USA, 2002.
12. Asimakopoulos, I.; Avendaño-Valencia, L.D.; Lützen, M.; Rytter, N.G.M. Data-driven condition monitoring of two-stroke marine diesel engine piston rings with machine learning. *Ships Offshore Struct.* **2023**, *19*, 1241–1253.
13. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *10*, P10008. [[CrossRef](#)]
14. Kim, D.; Antariksa, G.; Handayani, M.P.; Lee, S.; Lee, J. Explainable Anomaly Detection Framework for Maritime Main Engine Sensor Data. *Sensors* **2021**, *21*, 5200. [[CrossRef](#)] [[PubMed](#)]
15. Kankal, A.V.; Ghayal, S.D.; Hage, V.S.; Pawar, G.S.; Lodhe, S.P.; Kumar, M. AI-driven predictive maintenance in ocean energy systems. *Int. J. Sci. Res. Arch.* **2025**, *16*, 381–391. [[CrossRef](#)]
16. Kalafatelis, A.S.; Levis, G.; Giannopoulos, A.; Tsoulakos, N.; Trakadas, P. Explainable Predictive Maintenance of Marine Engines Using a Hybrid BiLSTM-Attention-Kolmogorov Arnold Network. *J. Mar. Sci. Eng.* **2026**, *14*, 32. [[CrossRef](#)]
17. Chaval, V.K.; Dharmavaram, V.G. AI-Driven Predictive Maintenance in Shipyards: Enhancing Project Management Efficiency and Operational Cost Reduction through Statistical, Data-Driven Strategies for MRO Services. *J. Inf. Syst. Eng. Manag.* **2025**, *10*, 63–75. [[CrossRef](#)]
18. Simion, D.; Postolache, F.; Fleacă, B.; Fleacă, E. AI-Driven Predictive Maintenance in Modern Maritime Transport—Enhancing Operational Efficiency and Reliability. *Appl. Sci.* **2024**, *14*, 9439. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.