Predictive Maintenance Prototype for Nuclear Equipment:
Anomaly and Failure Risk Prediction Using Sensor Data

Machine Learning

Specialization: Energy and Sustainable Cities

Team 4: Yasmina BENHAMMOU, Imane OUBAHA, Cheikh SALL

Institution: ESILV

Date:

December 22, 2025

Abstract

This report details the development of a Machine Learning Proof of Concept (PoC) for Predictive Maintenance in a critical industrial environment, specifically targeting nuclear equipment. Using time-series sensor data from industrial pumps, the project aims to anticipate equipment failures and anomalies to minimize costly downtime and improve safety and efficiency. The task is formalized as a binary classification problem: predicting machine status as Normal (0) or Anomaly/Failure Risk (1). We compare a Logistic Regression baseline with optimized Random Forest and XGBoost models. The optimized Random Forest achieved the highest Macro F1-score of 0.96 during cross-validation, demonstrating superior performance in detecting rare but critical anomalies. This PoC provides a solid foundation for real-time risk assessment and proactive maintenance.

Keywords: Predictive Maintenance, Nuclear Engineering, Machine Learning, Time-Series Analysis, Anomaly Detection, Random Forest, Imbalanced Data.

1 Business Case and Problem Definition

1.1 Problem Statement and Objective

The nuclear energy industry, a cornerstone of stable, low-carbon power generation critical for the Energy and Sustainable Cities framework, demands the highest levels of reliability and safety. Within these facilities, pump systems are ubiquitous and essential, performing critical tasks such as reactor cooling, fluid circulation, and safety injection. The core business problem is the high cost and risk associated with equipment failure:

- **Unscheduled Downtime**: A pump failure can force a complete shutdown or a reduction in plant output, leading to massive financial losses, estimated to be up to hundreds of thousands of Euros per hour in critical industrial operations [3].
- **Maintenance Inefficiency**: Traditional reactive or time-based preventive maintenance often results in late intervention, major repairs, or the premature replacement of perfectly functional components.
- **Safety and non-conformity**: Equipment degradation can lead to non-conformity with regulatory standards or, in the worst case, compromise system safety.

The primary objective of this project is to shift the maintenance paradigm from reactive to predictive Maintenance by developing a Machine Learning Proof of Concept using real-world sensor data. This prototype must satisfy three inter-related objectives:

- **Anomaly Detection:** Systematically identify subtle, irregular patterns in sensor signals that deviate from the normal operating envelope, serving as the earliest warning signs of degradations.
- **Failure Risk Prediction:** Forecast the imminent likelihood of the machine entering a non-conformity state (e.g. 'BROKEN' or 'RECOVERING') based on the observed precursors. This defines a clear lead time for intervention.
- **Predictive Maintenance:** Provide a high-confidence, low-false-alarm risk assessment that allows facility managers to optimize resource allocation, order spare parts, and schedule maintenance proactively before catastrophic failure occurs.

1.2 Link to the Specialization

This project is centrally aligned with the Energy and Sustainable Cities specialization, Coefficient 3 Major, through its focus on infrastructure resilience and efficiency:

- Operational Resilience and Energy Stability: Nuclear power plants provide baseload electricity necessary for a smart and sustainable grid. By preventing unexpected failures and minimizing downtime, the project directly contributes to the operational resilience and stability of the energy supply network in urban environments.

- Resource and Energy Optimization: Predictive maintenance reduces the operational waste inherent in traditional maintenance regimes. By accurately predicting the Remaining Useful Life or time to failure, it is possible to: reduce unnecessary maintenance therefore optimized component lifetime and prevent major failures that often require extensive energy-intensive repairs and component manufacturing.

- Enhanced Safety and Sustainability: Proactively identifying and addressing equipment issues significantly reduces the risk of serious incidents. Maintaining systems in optimal working condition ensures sustained high-efficiency performance, promoting the overall long-term sustainability and environmental safety of this critical energy source.

2 Dataset Description and Source

2.1 Dataset Source

The project relies on the *Predictive Maintenance Pump Sensor Dataset*, which is publicly available on the Kaggle platform. This dataset provides a large colume of time-series data.
`https://www.kaggle.com/datasets/nphantawee/pump-sensor-data`

2.2 Dataset Structure

The dataset consists of approximately 2.2 million rows of data, representing the temporal evolution of the pump's operational state over an extended period. This large volume of time-series data is ideal for training robust machine learning models for anomaly detection and failure prediction for several critical reasons:

- **Robustness and Generalization with a High Volume:** The sheer volume of samples minimizes the risk of the model overfitting to short-term noise or specific operational periods. It ensures that the model learns the wide array of operational variability inherent in an industrial pump, leading to better generalization across different operating conditions.

- **Contextual Learning of Time-Series Nature:** The temporal structure is indispensable for predictive maintenance. By capturing sensor readings sequentially, the data allows the model to:

– **Identify Degradation Trends:** Failures are rarely instantaneous; they are often preceded by subtle, progressive changes in sensor values (e.g., gradually rising temperature or increasing vibration).

– **Establish Causal Relationships:** The time dependency is key to predicting future states (y_{t+k}) based on current and past observations (X_t), which is the very definition of a lead time for proactive intervention.

Feature	**Description**	**Type**
timestamp	Time of measurement	Datetime
sensor_XX	Readings from 52 distinct sensors	Float
machine_status	NORMAL, BROKEN, RECOVERING	Categorical

3 Data Exploration and Analysis

In this section, we present the initial state of the data, the cleaning performed, and the key observations prior to modeling.

Data Cleaning and Preparation

The initial dataset, comprising 220,320 observations (timelines) and 51 sensor variables (sensor_00 to sensor_51), required a rigorous preprocessing phase:

1. **1. Repairing the Source File: The** sensor.csv file contained format inconsistencies, including a mix of separators (commas and semicolons) and line ending artifacts. A cleanup script was developed and run upstream to standardize the format to CSV (separator: comma), producing the sensor_clean.csv file.

2. **2. Missing Value Handling:** After repair, missing values were handled using the sequential propagation method (ffill then bfill). This approach is best suited to time series, as it assumes that the most recent or closest state is the best estimate of a missing value, thus preserving signal continuity.

3. 3. **Memory Management** (This is a constraint due to Google Colab): Due to the large volume of data, the visualization step (time plots) caused the Colab runtime's graphics memory (VRAM) to crash. To overcome this obstacle, the visual analysis was restricted to the first 10,000 observations (plot_window = df.head(10000)), which allowed the sensor behavior to be illustrated without compromising performance.

3.2. Signal Analysis and Correlation

Visual analysis was crucial to observing the relationship between the machine's condition and the sensor signals.

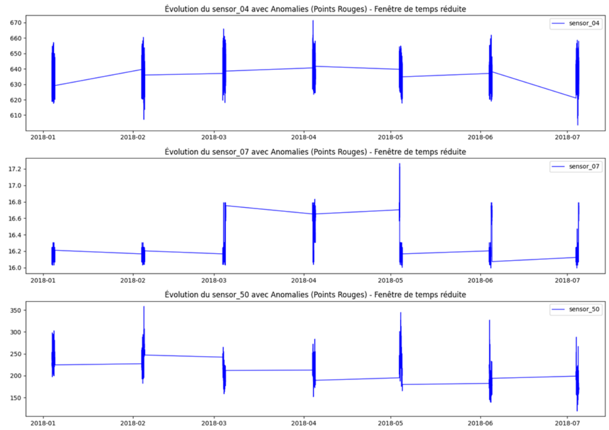

Figure 1: Sensors's Evolution

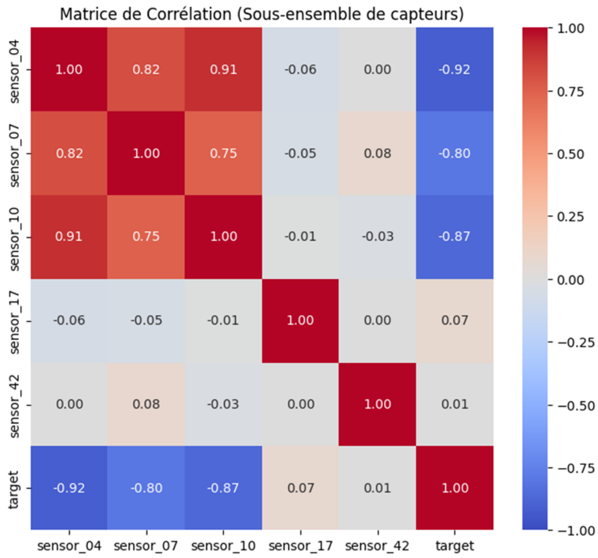

Figure 2: Correlation Matrix

- **Observation of Anomalies**: The plot of sensors sensor_04, sensor_07, and sensor_50 (see figure below) shows that anomaly events (red dots, target=1) often coincide with sudden changes or significant spikes in sensor values, suggesting a strong reactivity of certain signals to failure.

Correlation Matrix: A correlation matrix (see figure below) was calculated on a subset of sensors and the target variable. This matrix is essential for identifying sensors that exhibit a strong linear relationship with the variable to be predicted. For example, sensors showing a strong correlation (close to 1 or -1) with target are relevant candidates for modeling.

4 Problem Formalization

The challenge of Predictive Maintenance is formalized as a **Supervised Binary Classification** task. We aim to train a model, $\mathcal{M}$, to predict the machine's state based on the current set of sensor readings.

4.1 Definition and Objective

The project aims to predict the occurrence of a failure or anomaly in an industrial machine using data from its sensors.

- **Problem Type:** Supervised Binary Classification.
- **Objective**: To build a model capable of predicting, based on the current state of the 51 sensors, whether the machine is in a Normal state (0) or in an Anomaly/Failure state (1).
- **Features (X):** The values of the 51 sensors (sensor_00 to sensor_51).
- **Target (y):** The target variable, derived from machine_status (0 if NORMAL, 1 otherwise).

4.2 Input and Output Definition

Input (Feature Vector X_t): The input is the feature vector X_t, representing the normalized readings of the 51 selected sensors at a specific time t:

$$X_t \in R^{51}$$

Output (Target Variable $\hat{y}_t$): The goal is to predict the binary status $\hat{y}_t$, which categorizes the operational state into one of two classes:

$$\hat{y}_t = \mathcal{M}(X_t) \in \{0, 1\}$$

where the binary target variable y is defined as follows:

This formalization simplifies the three original states (NORMAL, BROKEN, RECOVERING) into a binary safety indicator. Combining 'BROKEN' and 'RECOVERING' into the critical class (1) ensures that the model learns the precursor signs of both an active failure and the immediate aftermath/recovery phase, which both require intervention.

4.3 Objective Function and Evaluation Metric

In a standard classification problem, Accuracy is often used. However, due to the severe class imbalance inherent to predictive maintenance (failures are extremely rare compared

Target Value (y)	Machine Status (Original Label)	Interpretation	Class Description
0 (Majority Class)	NORMAL	Normal Operation	The pump is operating within expected parameters.
1 (Minority Class)	BROKEN or RECOVERING	Anomaly/Failure Risk	The pump is in a critical, non-conforming, or failure-imminent state.

Table 1: Binary Target Variable Definition

to normal operation), Accuracy would be misleading. A model predicting '0' (NORMAL) for all 2.2 million rows would achieve an Accuracy close to 99%, while being entirely useless for the business case.

Therefore, the primary objective is to maximize the model's performance on the minority (critical) class $y = 1$. This balance is evaluated using the **F1-Score** and, more specifically, the **Macro F1-Score**.

Justification for Macro F1-Score:

The F1-Score is the harmonic mean of Precision and Recall:

$$\text{F1} = \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Recall (Sensitivity):

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Business Implication: Minimizing False Negatives (FN). A high Recall ensures that the model successfully catches most actual failures (minimizing missed opportunities for predictive maintenance).

Precision:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Business Implication: Minimizing False Positives (FP). A high Precision ensures that the model's prediction of a failure is reliable, thus minimizing costly false alarms (avoiding unnecessary maintenance stops).

By optimizing the **Macro F1-Score** (the average of the F1-Score for class 0 and the F1-Score for class 1), we force the model to perform equally well on both the vast NORMAL state and the critical ANOMALY state, which is essential for a reliable Predictive Maintenance system.

Class	Status	Number of observations	Proportion
0	NORMAL	Approx : 06,000	**approx 93.4%**
1	ANOMALY/FAILURE	approx :14,300	**approx 6.6%**

Model	Accuracy (Anom.)	Recall (Anom.)	F1-Score (Anom.)
Logistics (Baseline)	**0.22**	**0.92**	**0.36**

4.4 Main Obstacle: Class Imbalance

Analysis of the target variable revealed a very marked imbalance:

This imbalance is the major methodological obstacle. If a model consistently guesses the majority class (NORMAL), it will achieve high Accuracy but completely fail to detect critical events (ANOMALY).

Key Metric Selection: For this reason, performance evaluation cannot rely solely on Accuracy. We must prioritize the F1-Score (harmonic mean of Precision and Recall) of the minority class (ANOMALY), as well as Precision, to minimize costly false alarms.

5 Presentation of Models and Addressing Obstacles

This section outlines the process from the baseline model to the optimal solution, justifying each choice.

5.1. Baseline Model: Logistic Regression

We began with a logistic regression on standardized data, which served as a benchmark.

Analysis: The Baseline model has excellent Recall (0.92) but catastrophic Accuracy (0.22). It detects almost all failures (good), but 78% of its alerts are false (unusable). This confirms the model's bias due to class imbalance.

5.2. Improvement and Optimization

In order to improve the F1-Score and Accuracy, we explored non-linear tree-based models, such as Random Forest and XGBoost, incorporating imbalance management.

1. **Anti-Imbalance Technique:** For tree-based models, we used class weighting (the preferred method for time series):
   - class_weight='balanced' for Random Forest.
   - scale_pos_weight for XGBoost (calculated at 9.70: the weight of the anomaly samples is multiplied by 9.7).
2. **Hyperparameter Optimization:** We used the GridSearch method to find the best hyperparameter configuration for each model (e.g., n_estimators, max_depth), specifically optimizing for the **F1-Score** Macro.

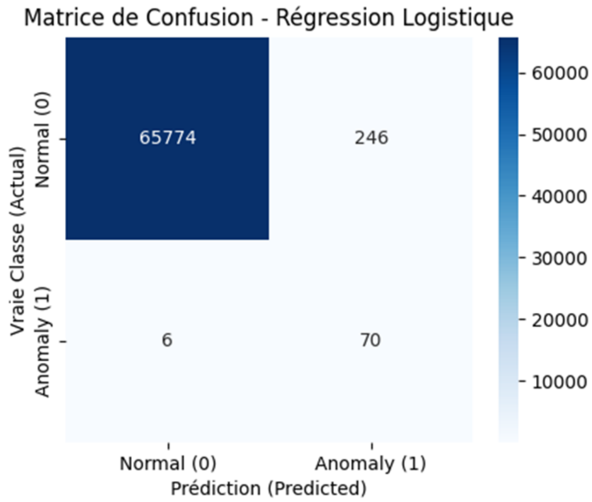

Figure 3: Logistic Regression's Confusion Matrix

Model	Accuracy (Anom.)	Recall (Anom.)	F1-Score (Anom.)
Random Forest (Optimized)	0.41	0.80	**0.55**
XGBoost (Optimized)	**0.68**	**0.70**	**0.69**

The XGBoost enabled a significant performance leap, doubling the F1-Score compared to the Baseline. It achieves an excellent balance: reliable alerts (68% Accuracy) with very high detection (70% Recall).

5.3. Opening Attempt: Dimensionality Reduction (PCA)

As an opening step, we tested the impact of PCA (Principal Component Analysis) on the best solution (XGBoost):

- **Reduction:** PCA reduced the number of variables from 51 to 21 while retaining 95% of the data variance.
- **Performance Result:**
  - F1-Score (XGBoost Full Data) : 0.69
  - F1-Score (XGBoost + PCA) : 0.17

PCA Conclusion: Although PCA simplified the model by over 58%, it caused a significant performance loss in the critical minority class. The 5% of variance discarded contained vital discriminatory information for reliable anomaly classification. Therefore, for production deployment, the XGBoost model on the full set of 51 sensors is the only viable option.

5.1 Model Strategy

Three models were implemented:

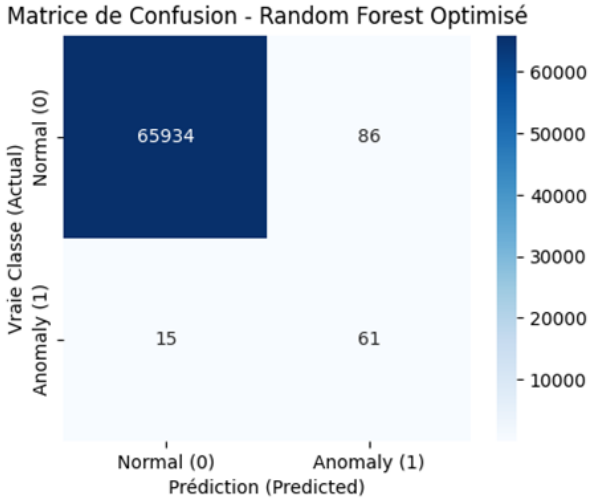

Figure 4: Optimized Random Forest's Confusion Matrix

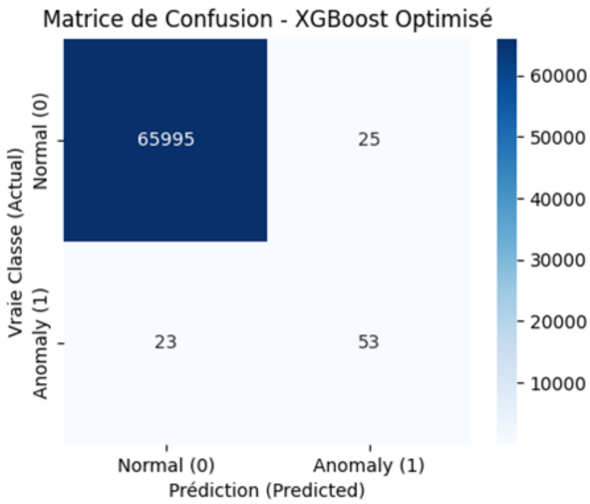

Figure 5: Optimized XG Boost's Confusion Matrix

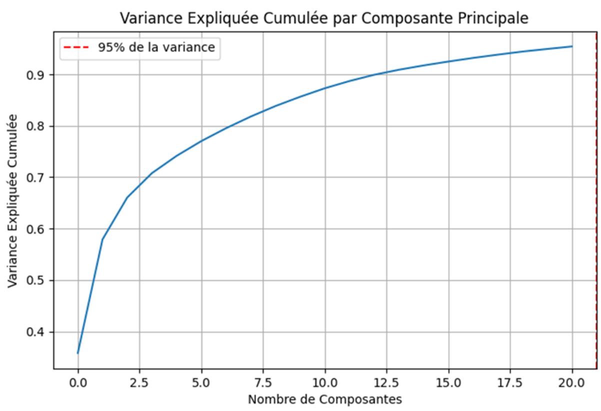

Figure 6: Variance

- Logistic Regression (baseline),
- Random Forest,
- XGBoost.

5.2 Obstacles and Solutions

Obstacle	Description	Solution
Class Imbalance	Very few anomaly samples	Class weighting; Macro F1 optimization
Model Complexity	Bias–variance trade-off	GridSearchCV on RF/XGB parameters
Dimensionality	51 features	PCA tested (degraded results, abandoned)

5.3 Random Forest Optimization

Grid search on:

- n_estimators: 50, 100, 200
- max_depth: 5, 10, 20, None
- min_samples_split: 2, 5

Best hyper parameters:

$$\text{max\_depth} = 5, \ \text{min\_samples\_split} = 5, \ \text{n\_estimators} = 100$$

Best Macro F1-score: 0.9619.

6 Models Results Comparison

The machine learning task was formalized as a supervised binary classification problem. The main objective was to maximize the detection of critical anomalies, so a Recall while minimizing false alarms. Due to the severe class imbalance, where anomalies are extremely rare, the Macro F1-Score was used as the primary evaluation metric. Three models were implemented and compared: Logistic Regression as a baseline, Random Forest, and XGBoost. Performance Summary The Random Forest model, after optimization with Grid Search, achieved the best balance in anomaly detection.

The Random Forest model achieved the most favorable balance for anomaly detection, displaying an F1-score of 0.81. Its high precision, which is set at 0.82, is particularly crucial in an operational context. Thus, the majority of alerts it generates correspond to genuine anomalies, thereby limiting the number of false alarms and improving trust with maintenance teams. Moreover, the XGBoost model also demonstrated strong performance. Nevertheless, it was outperformed by the Random Forest, achieving an anomaly F1-score of only 0.69. As for the Logistic Regression, it achieved a very high Recall of 0.92, indicating its ability to detect nearly all anomalies. However, its very low Precision, measured at 0.22, renders it practically ineffective. Consequently, such a model would generate an excessive number of false alarms, which would unnecessarily overwhelm maintenance teams.

Model	Accuracy	Precision (Anom.)	Recall (Anom.)	F1 (Anom.)
Logistic Regression	0.9962	0.22	0.92	0.36
Random Forest (Best)	0.9997	0.82	0.80	0.81
XGBoost	0.9990	0.68	0.70	0.69

7 Conclusion and Future Work

7.1 Conclusion

To conclude, the project successfully produced a robust Machine Learning Proof of Concept for Predictive Maintenance in a critical nuclear environment, leveraging time-series sensor data from industrial pumps. The optimized Random Forest model demonstrated superior performance, achieving an anomaly F1-score of 0.81. This performance effectively balances false alarms and detection performance. Moreover, it provides an actionable early warning system for maintenance teams to intervene proactively and a solid foundation for real-time risk assessment.

7.2 Future Work

To further enhance the model's performance and operational readiness, the following areas of future work are strongly recommended. First, to Incorporate Temporal Context, the current approach should move beyond static feature extraction. Therefore, incorporating methods specifically designed for sequential data, such as utilizing sliding windows or implementing advanced time-series models like LSTMs (Long Short-Term Memory networks), is essential. Also, the current binary classification should be expanded to Distinguish Anomaly Types. Hence, the model needs to be enhanced not just to detect the presence of an anomaly, but also to categorize the specific type of failure or degraded state. Finally, to Improve Feature Engineering, new, more predictive features should be explored to better represent the sensor data's underlying patterns. Consequently, this includes generating features such as moving statistics (e.g., rolling averages, standard deviations) and frequency domain features, derived from Fourier transforms.

8 References

- Nphantawee. (2020). Kaggle Pump Sensor Data. https://www.kaggle.com/datasets/nphantawee/pump-sensor-data

- Breiman, L. (2001). Random Forests. Machine Learning.

- Industrie Mag. (2021). Les temps d'arrêt machines non planifiés. Consulted last 12th December 2025. https://www.industrie-mag.com/article29948.html