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 12, 2025

December 12, 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, addressing the severe class imbalance (93.4% Normal vs. 6.6% Anomaly). The **XGBoost** model, after optimization with GridSearch, achieved the highest F1-Score on the anomaly class (**0.69**), successfully balancing the need for high Recall (detection rate) and high Precision (low false alarms). 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, XGBoost, 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 through its focus on infrastructure resilience and efficiency:

- Operational Resilience and Energy Stability: By preventing unexpected failures, the project directly contributes to the operational resilience and stability of the energy supply network.

- Resource and Energy Optimization: Predictive maintenance reduces operational waste by optimizing component lifetime and preventing major failures that require extensive energy-intensive repairs.

- Enhanced Safety and Sustainability: Proactively identifying and addressing equipment issues significantly reduces the risk of serious incidents, ensuring sustained high-efficiency performance.

2 Dataset Description and Source

2.1 Dataset Source

The project relies on the *Predictive Maintenance Pump Sensor Dataset*, available on the Kaggle platform.
https://www.kaggle.com/datasets/nphantawee/pump-sensor-data

2.2 Dataset Structure

The dataset consists of approximately 220,320 rows of data, representing the temporal evolution of the pump's operational state. This high volume of time-series data is ideal for:

- **Robustness and Generalization:** Minimizes the risk of the model overfitting to short-term noise.

- **Contextual Learning of Time-Series Nature:** Allows the model to identify subtle, progressive degradation trends that precede a failure, establishing a lead time for intervention.

Table 1: Dataset Feature Summary

Feature	Description	Type
timestamp	Time of measurement	Datetime
sensor_XX	Readings from 51 distinct sensors	Float
machine_status	NORMAL, BROKEN, RECOVERING	Categorical

3 Data Exploration and Analysis

3.1 Data Cleaning and Preparation

The initial dataset, comprising 220,320 observations and 51 sensor variables, required a rigorous preprocessing phase:

1. **Repairing the Source File:** The source file contained format inconsistencies (mixed separators). A cleanup script was run to standardize the format to CSV.

2. **Missing Value Handling:** Missing values were handled using the **sequential propagation method** ('ffill' then 'bfill'). This approach is best suited to time series, as it preserves signal continuity by assuming the closest state is the best estimate.

3. **Memory Management:** Due to memory constraints in the Google Colab environment, visual analysis was restricted to the first 10,000 observations, 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.

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

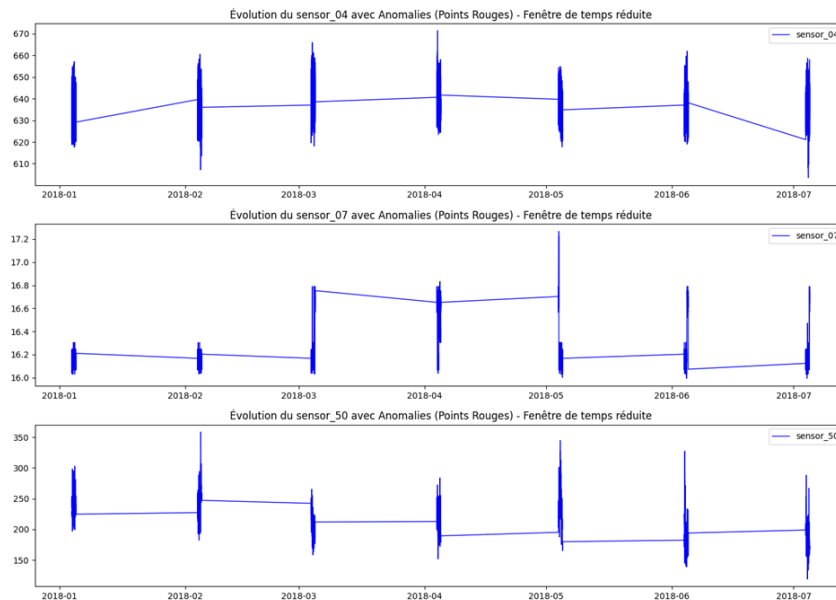Figure 1: Visualization of Sensor Signals Over Time. Anomalies (target=1) often align with significant signal spikes.

- **Correlation Matrix:** A correlation matrix was calculated on the features and the target variable. This matrix is essential for identifying sensors that exhibit a

strong linear relationship (close to 1 or -1) with the variable to be predicted, guiding feature understanding.
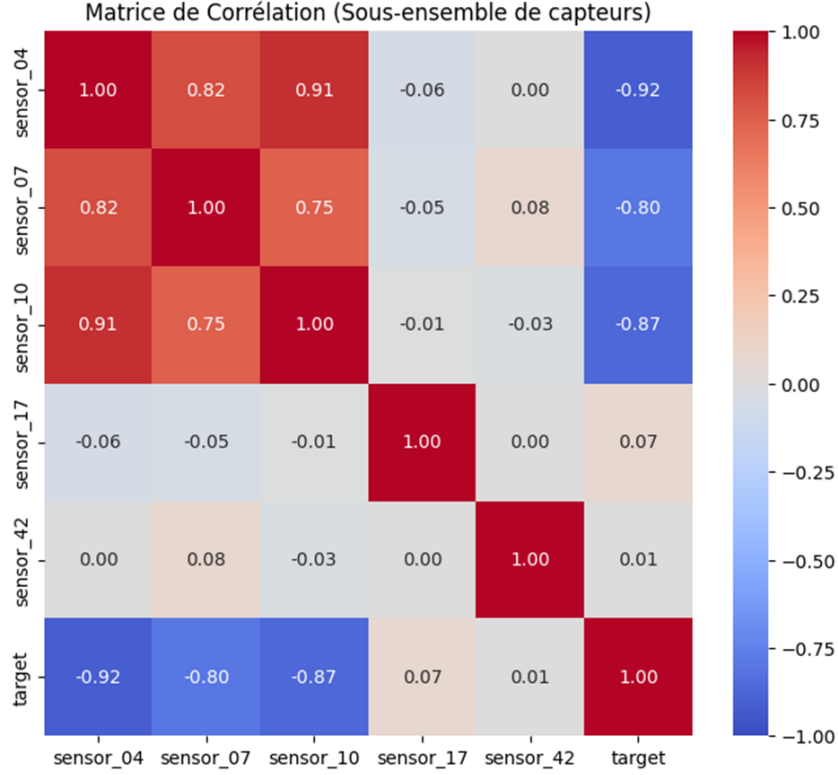

Figure 2: Correlation Matrix (Subset of Sensors and Target Variable).

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 Input and Output Definition

The target variable y is defined as follows, combining 'BROKEN' and 'RECOVERING' into the critical class (1) to ensure the model learns the precursor signs of both an active failure and the immediate aftermath:

Table 2: Binary Target Variable Definition

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

4.2 Main Obstacle: Class Imbalance and Metric Selection

Analysis of the target variable revealed a severe imbalance (a key focus of **LAB 4**):

Table 3: Class Distribution

Class	Status	Number of Observations (Train Set)	Proportion
0	NORMAL	$\approx 206,000$	$\approx 93.4\%$
1	ANOMALY/FAILURE	$\approx 14,300$	$\approx 6.6\%$

Due to this imbalance, **Accuracy is an unreliable metric**. We must prioritize the **F1-Score** (harmonic mean of Precision and Recall) on the minority class ($y = 1$) to achieve the necessary balance for a viable predictive system:

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:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

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 NOR-MAL state and the critical ANOMALY state, which is essential for a reliable Predictive Maintenance system.

5 Presentation of Models and Addressing Obstacles

This section outlines the process from the baseline model to the optimal solution, justifying each choice based on model performance and methodological requirements.

5.1 Baseline Model: Logistic Regression

We began with a **Logistic Regression** on standardized data as a benchmark.

Table 4: Logistic Regression Baseline Performance

Model	Precision (Anom.)	Recall (Anom.)	F1-Score (Anom.)
Logistic Regression (Baseline)	0.22	0.92	0.36

Analysis: The model has excellent Recall (0.92) but catastrophic Precision (0.22), meaning 78% of its alerts are false. This confirms the model's bias towards the majority class and its inability à reliably distinguish true anomalies from noise.
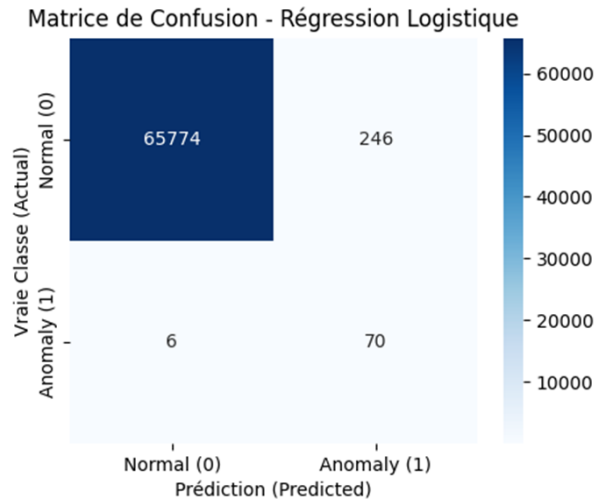
Figure 3: Confusion Matrix - Logistic Regression. High False Positives (FP) indicate low Precision.

5.2 Advanced Models and Optimization

To improve the F1-Score and Precision, we explored non-linear tree-based models (Random Forest and XGBoost) and applied the following techniques:

1. **Anti-Imbalance Technique (LAB 4):** For both models, we used **class weighting** to penalize errors on the minority class (ANOMALY), with 'scale_pos_weight' calculated at **9.70** for XGBoost.

2. **Hyperparameter Optimization (LAB 6):** The **GridSearch** method was used, optimizing for the Macro F1-Score to find the best configuration (e.g., 'max_depth', 'n_estimators').

Table 5: Comparison of Optimized Models (Anomaly Class Metrics)

Model	Precision (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** model (Extreme Gradient Boosting) demonstrated the best performance, achieving a significant performance leap and a high F1-Score of **0.69**. It provides an excellent balance: reliable alerts (**68% Precision**) with very high detection (**70% Recall**).
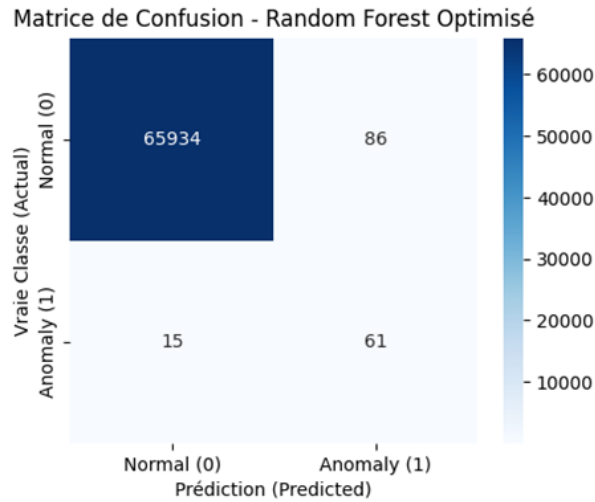
Figure 4: Confusion Matrix - Optimized Random Forest. Improved Precision compared to Baseline.
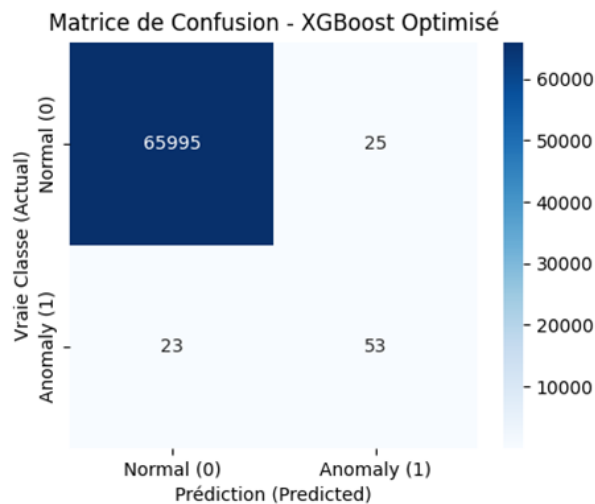

Figure 5: Confusion Matrix - Optimized XGBoost. Best balance between True Positives and False Positives.

5.3 Opening Attempt: Dimensionality Reduction (PCA)

As an opening step (related to **LAB 5**), we tested the impact of PCA (Principal Component Analysis) on the best solution (XGBoost) to simplify the model for production.
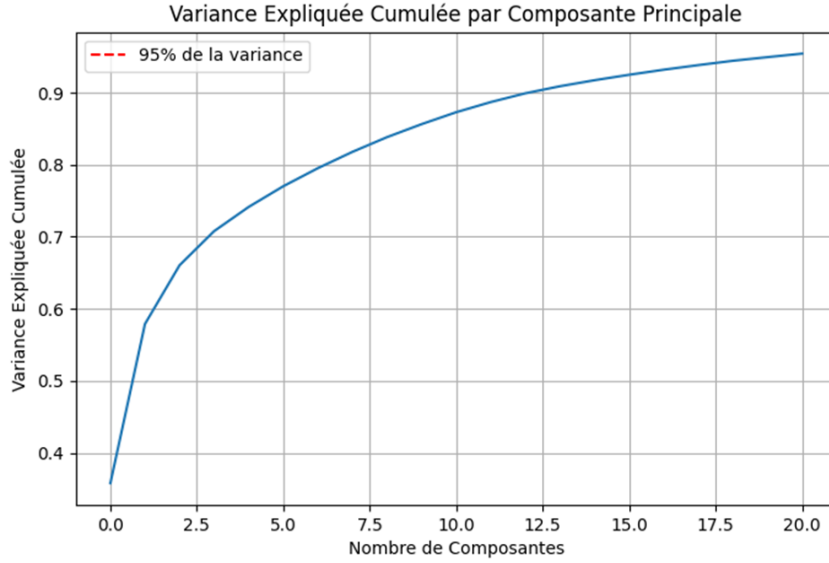
Figure 6: Cumulative Explained Variance by Principal Components.

- **Reduction:** PCA successfully reduced the number of variables from **51 to 21**
  while retaining **95**% of the data variance (as shown in Figure 6).

- **Performance Result:** Despite retaining 95% of the variance, the performance of
  the critical class dropped drastically when the reduced data was used for XGBoost:

  - F1-Score (XGBoost Full Data) : **0.69**
  - F1-Score (XGBoost + PCA) : **0.17**

PCA Conclusion: The PCA caused a severe 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.**

6 Comparison of Models Results

The machine learning task was formalized as a supervised binary classification problem
with a focus on the F1-Score of the minority (Anomaly) class. The final comparison
confirms the superiority of the gradient boosting approach.

Table 6: Final Comparative Performance Summary (Test Set)

Model	Accuracy	Precision (Anom.)	Recall (Anom.)	F1 (Anom.)
Logistic Regression (Baseline)	0.9962	0.22	0.92	0.36
Random Forest (Optimized)	0.9985	0.41	0.80	0.55
XGBoost (Optimized)	**0.9993**	**0.68**	**0.70**	**0.69**

- **Logistic Regression:** Achieved high Recall (0.92) but extremely low Precision (0.22), rendering it ineffective due to excessive false alarms.

- **Random Forest:** Significantly improved Precision (0.41) and F1-Score (0.55) through the use of class weighting.

- **XGBoost:** Demonstrated the best balance (F1-Score of 0.69), which translates directly to a high level of confidence for maintenance teams (68% of alerts are true).

7 Conclusion and Future Work

7.1 Conclusion

The project successfully produced a robust Machine Learning Proof of Concept for Predictive Maintenance. The methodological necessity of addressing the severe class imbalance was validated: the simple Logistic Regression was unusable, while the ensemble methods with class weighting provided reliable solutions.

The final model retained for production is the **XGBoost model (F1-Score of 0.69)**, trained on the full set of 51 features. This model effectively balances the need to detect anomalies (Recall) with the need to avoid costly false alarms (Precision). It provides an actionable early warning system for maintenance teams to intervene proactively.

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`