Explainable AI for Industrial Predictive Maintenance in the UNDERPIN Manufacturing Dataspace*

Nikos Kapetanas[1,*,†], Axel Weissenfeld[2,*,†], Harry Nakos[1,†], Yannis Stavrakas[1,†], Stelios Sartzetakis[1,†], Vasilis Gkolemis[1,†] and Theodore Dalamagas[1,†]

[1]*Information Management Systems Institute, Athena Research Center, Athens, Greece*
[2]*Austrian Institute of Technology, Vienna, Austria*

Abstract

Predictive maintenance in manufacturing relies heavily on machine learning models to detect anomalies and predict failures before they occur. However, the complexity of these models often limits interpretability, which reduces trust and hinders adoption in industrial contexts. This has raised the demand for explainable AI (XAI) techniques. This paper presents a real-world implementation of Effector, an explainable AI package for tabular data, in an industrial predictive maintenance system within a Dataspace environment, to interpret predictions from a CatBoost model trained on time series manufacturing data for anomaly detection. Unlike traditional "black box" AI solutions, Effector provides transparent decision-making through global and regional effect plots that reveal how each feature influences model predictions. We demonstrate how Effector's regional explanations reduce heterogeneity in feature effects, improving interpretability and providing actionable insights for maintenance decision-makers. The integration of Effector XAI within a manufacturing Dataspace [1] highlights the potential of semantic technologies to enhance transparency and trust in industrial AI systems.

Keywords

Explainable AI, Predictive Maintenance, Manufacturing Dataspace, CatBoost, Industrial IoT, Anomaly Detection

1. Introduction

The increasing adoption of machine learning (ML) in industrial predictive maintenance has revolutionized how manufacturing companies approach equipment reliability and operational efficiency. Predictive maintenance offers a proactive approach to identify potential issues before they escalate, reducing downtime and maintenance costs while improving operational efficiency. Machine learning models like CatBoost have shown strong performance in anomaly detection on time series sensor data. However, their "black box" nature presents a significant challenge and often limits user trust, hindering adoption in industrial applications where equipment failures can result in high costs, lost production, and potential safety issues.

Explainable AI has emerged as a critical solution to address the transparency challenges posed by increasingly sophisticated machine learning models. In industrial predictive maintenance, the ability to understand and validate model predictions is not merely desirable — it is essential for responsible deployment. The complex relationships between operational parameters, environmental conditions, and equipment health require explanations that can reveal not just what the model predicts, but why specific decisions are made under different operational contexts.

This paper addresses the challenge of deploying trustworthy AI in industrial predictive maintenance by integrating Effector, an explainable AI library, within a manufacturing Dataspace environment. Our work focuses on wind turbine predictive maintenance, where complex sensor interactions and

---

*This paper demonstrates the practical application of Effector XAI in industrial predictive maintenance systems for enhanced transparency and trust.

*Corresponding author.
†These authors contributed equally.
nkapetanas@athenarc.gr (N. Kapetanas); axel.weissenfeld@ait.ac.at (A. Weissenfeld); xnakos@athenarc.gr (H. Nakos); ys@athenarc.gr (Y. Stavrakas); stelios@athenarc.gr (S. Sartzetakis); vgkolemis@athenarc.gr (V. Gkolemis); dalamag@athenarc.gr (T. Dalamagas)

varying operational conditions make model interpretability crucial for maintenance decision-making. By embedding Effector's XAI capabilities within the Dataspace infrastructure, we demonstrate how transparent AI can be delivered as a service to multiple users.

Contributions:

- Implementation of Effector XAI in a real-world industrial predictive maintenance system for wind turbine anomaly detection.
- Integration of explainable AI capabilities within a manufacturing Dataspace architecture enabling transparent AI-driven maintenance decisions.
- Demonstration of how regional explanations reveal critical operational thresholds and individual equipment behavioral patterns.
- Practical insights into how XAI enhances maintenance strategy optimization through identified efficiency zones and stress conditions.

2. Background and Related Work

2.1. Predictive Maintenance and Machine Learning

Predictive maintenance has evolved from traditional time-based and condition-based approaches to sophisticated data-driven methodologies leveraging machine learning algorithms. The "black-box" nature of these models created the need to apply Explainable Artificial Intelligence methods. The two main types of approach are model-agnostic and model-specific. Model-agnostic techniques are independent of the underlying model architecture, making them widely applicable for interpreting diverse machine learning models [2]. Examples include feature importance methods such as Shapley values [3] and surrogate model approaches like LIME [4]. In contrast, model-specific approaches exploit internal model structures for interpretability.

2.2. XAI Techniques and Explanations in Industrial Applications

XAI techniques can provide local explanations (for individual predictions) or global explanations (for entire datasets) [2]. LIME [4] and SHAP [5] are widely used for local interpretability. The adoption of XAI in industrial contexts has gained significant attention as organizations seek to balance model performance with interpretability requirements. Lundberg and Lee [5] introduced SHAP (SHapley Additive exPlanations), which has become a standard approach for feature importance analysis in industrial applications. However, SHAP's focus on individual predictions limits its ability to provide insights into global model behavior across different operational conditions.

Partial Dependence Plots (PDP) [6] and Individual Conditional Expectation (ICE) [7] plots have emerged as powerful tools for understanding feature effects, but their application in complex industrial settings with heterogeneous operational conditions remains challenging.

2.3. XAI Dashboards and Visualization Systems for Predictive Maintenance

In the context of predictive maintenance, recent works have focused on the operationalization of XAI in predictive maintenance. Marmara et al. [8] proposed an explainable AI model for predictive maintenance and spare parts optimization, using LIME to interpret model outputs and provide actionable insights for maintenance planning. Ameli et al. [9] leveraged XAI to localize sensor anomalies in glass production, and Senoner et al. [10] used SHAP to uncover relationships between production parameters and process quality.

Recent surveys and frameworks further emphasize the importance of XAI dashboards for operational excellence in predictive maintenance. Cummins et al. [11] provided a systematic review of XAI methods in predictive maintenance, highlighting current challenges and opportunities. The Politecnico di Milano framework [12] proposed a user-friendly dashboard that translates complex XAI insights into actionable guidance for maintenance teams, bridging the gap between technical experts and operational users.

These works collectively demonstrate the growing maturity of XAI in predictive maintenance, particularly the role of interactive dashboards in making AI-driven insights transparent, actionable, and accessible for industrial practitioners.

2.4. Effector XAI

Effector is a Python library dedicated to generating global and regional feature effect explanations for tabular data models [13]. It is model-agnostic and integrates easily with popular ML frameworks. The key innovation of this library lies in its ability to automatically detect subspaces — defined by logical rules on features — where regional effects exhibit reduced heterogeneity compared to global effects. The library implements several key innovations: regional variants of major explanation methods including PDP, ALE, and the novel Robust and Heterogeneity-aware ALE (RHALE); sophisticated algorithms for automatic subspace detection that minimize heterogeneity while maintaining interpretability; and a unified API that facilitates comparison between different explanation methods.

Partial Dependence Plots reveal how features influence model predictions by showing the average effect of a feature across all possible values while marginalizing over all other features. The key insight behind PDP is that by averaging predictions across the dataset while varying only the feature of interest, we can isolate that feature's contribution to the model's decision-making process. This approach creates a curve that shows whether increasing or decreasing a feature value tends to increase or decrease the predicted outcome on average.

However, PDP has limitations when feature effects are heterogeneous across different subgroups in the data. Individual Conditional Expectation (ICE) plots address this limitation by showing how each individual instance's prediction changes as a single feature varies, creating multiple curves that reveal the variability hidden in PDP's averaging process. When ICE curves diverge significantly, it indicates that the feature's effect varies considerably across different instances or operational conditions.

2.5. Differences from Existing Approaches

Our work differs from existing XAI implementations in industrial predictive maintenance in several key aspects. First, we integrate Effector's regional explanation capabilities within a manufacturing Dataspace environment, enabling context-aware explanations that adapt to different operational conditions. Second, our approach combines global model understanding through PDP with individual equipment behavioral analysis through ICE plots, providing both fleet-wide insights and equipment-specific recommendations. Third, we implement an asynchronous processing architecture that handles computationally intensive XAI analysis while maintaining real-time operational capabilities. Finally, our system provides actionable maintenance insights by identifying specific operational thresholds and efficiency zones rather than generic feature importance rankings.

2.6. UNDERPIN Manufacturing Dataspace

The UNDERPIN Manufacturing Dataspace focuses on dynamic asset management and predictive maintenance for refineries and wind farms. Modern refineries generate massive amounts of operational data every day from sensors monitoring pressure, temperature, vibration, and more. But the real value comes not just from collecting data, but from turning it into actionable insight. On UNDERPIN the the focus is on using machine learning to anticipate equipment failures before they happen. Instead of relying on fixed maintenance schedules or reacting to unexpected breakdowns, the system continuously analyzes sensor data to detect early signs of wear, degradation, or abnormal behavior across the entire production chain. Within Wind Turbine Generators (WTGs), sensitive electrical systems and critical mechanical components installed inside the nacelle and the tower are prone to failures. While electrical failures are more frequent, mechanical failures, such as those in the gearbox, bearings, shafts, yaw, and pitch systems, are costly and have a significant impact on WTG performance, leading to prolonged downtimes, high maintenance costs, and production losses. The result is a refinery operation that becomes smarter over time, with better resilience, fewer interruptions, and more informed decision-making.

3. Methodology

3.1. Predictive Maintenance Use Case

Our use case focuses on a manufacturing Dataspace [1] containing a year-long dataset of SCADA sensor measurements, recorded at 10-minute intervals, from a single wind turbine generator (WTG) within wind farm multivariate time series sensor data from industrial equipment. The goal was to predict anomalies indicating potential failures. A CatBoost classifier trained for multivariate time series forecasting was used on historical data from 2019-01-01 to 2019-09-30 for training and from 2019-10-01 to 2019-12-28 for testing, achieving high accuracy predictions by detecting temperature anomalies. Data preprocessing involved correcting timestamp errors and out-of-range values, while removing erroneous entries that could not be corrected. The dataset was filtered to exclude non-operational conditions like wind speeds outside the 5-25 m/s range, blade pitch angles above 40°, and periods of restricted power or grid constraints. This filtering ensured the training data represented normal wind turbine generator operating conditions. Key health-indicative features were selected based on their predictability from ambient conditions and historical measurements—for example, generator bearing temperature serves as a reliable indicator of generator health status.

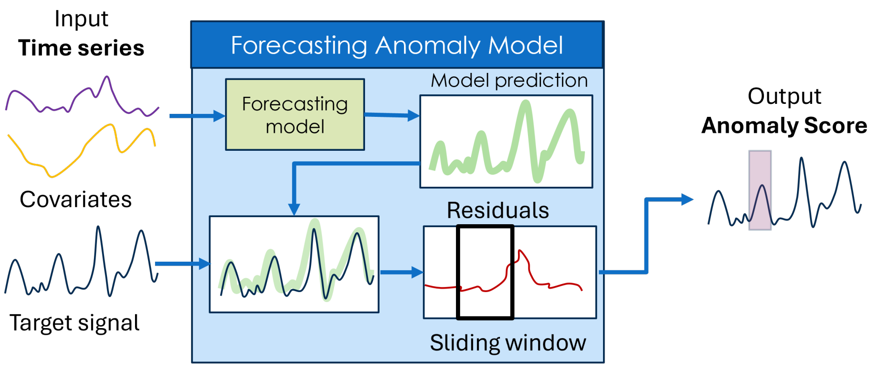

Figure 1: Forecasting on UNDERPIN

The covariates used to make these temperature predictions include:

- Generator RPM Max
- Nacelle Temperature Average
- Generator Cooling Water Temperature Average
- Rotation On/Off Status
- Missing Data Flags
- Active power generation from two sources

The target variable is Generator Bearing Temperature Average. The dataset contains 12,672 instances.

3.2. Applying Effector for Explainability

The core XAI functionality is implemented through the Effector library, providing model-agnostic explanations via Partial Dependence Plots (PDP) and Individual Conditional Expectation (ICE) plots. In our wind turbine use case, PDP analysis reveals the average effect of each operational parameter on bearing temperature predictions, while ICE plots expose individual turbine behavioral variations that would be hidden in global averages.

The implementation process involves three key steps: first, loading the trained CatBoost model and preprocessed SCADA data; second, generating PDP curves for each feature across its operational range; and third, overlaying ICE curves to visualize prediction heterogeneity across individual turbine instances. This combination provides maintenance engineers with both fleet-wide operational insights and equipment-specific behavioral patterns, enabling targeted maintenance strategies based on individual turbine characteristics rather than generic industry averages.

3.3. Integration in Manufacturing Dataspace

We developed an Effector-as-a-Service application that was integrated into a manufacturing Dataspace platform, enabling maintenance engineers to explore model behavior interactively. The application accepts data from the Dataspace, which is located on a MinIO bucket, along with start and end date parameters to apply Effector analysis on the time series data. The integration architecture provides an interface between the predictive models and the explanation system, allowing real-time interpretation of predictions within the existing industrial workflow.

The architecture shown in Figure 2 demonstrates how sensor data flows through the predictive maintenance pipeline with integrated explainability.

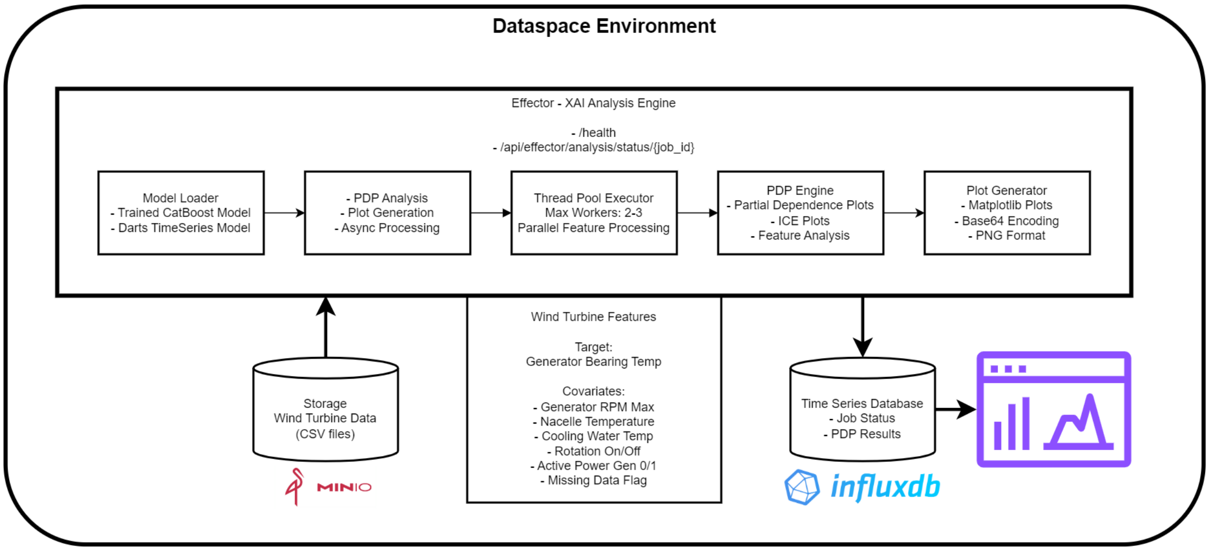

Figure 2: Integration architecture of Effector XAI within the manufacturing Dataspace environment

To handle computationally intensive XAI analysis, the system implements an asynchronous processing architecture where each analysis request generates a unique identifier (UUID) for tracking purposes. Jobs progress through defined states: started ⟶ processing ⟶ completed/failed. Feature-level parallelization uses a ThreadPoolExecutor, enabling concurrent PDP generation while preventing resource exhaustion, with real-time progress tracking, providing transparency for long-running analyses.

Matplotlib-based visualizations combine PDP curves with ICE plots, showing both average effects and individual variability. The generated plots are encoded as base64 strings and stored in InfluxDB with comprehensive metadata including feature names, target variables, timestamps, and job identifiers for reproducibility and audit trails. Status updates are persisted to InfluxDB for monitoring dashboard integration, enabling maintenance engineers to track the analysis progress and access the results through the Dataspace platform interface.

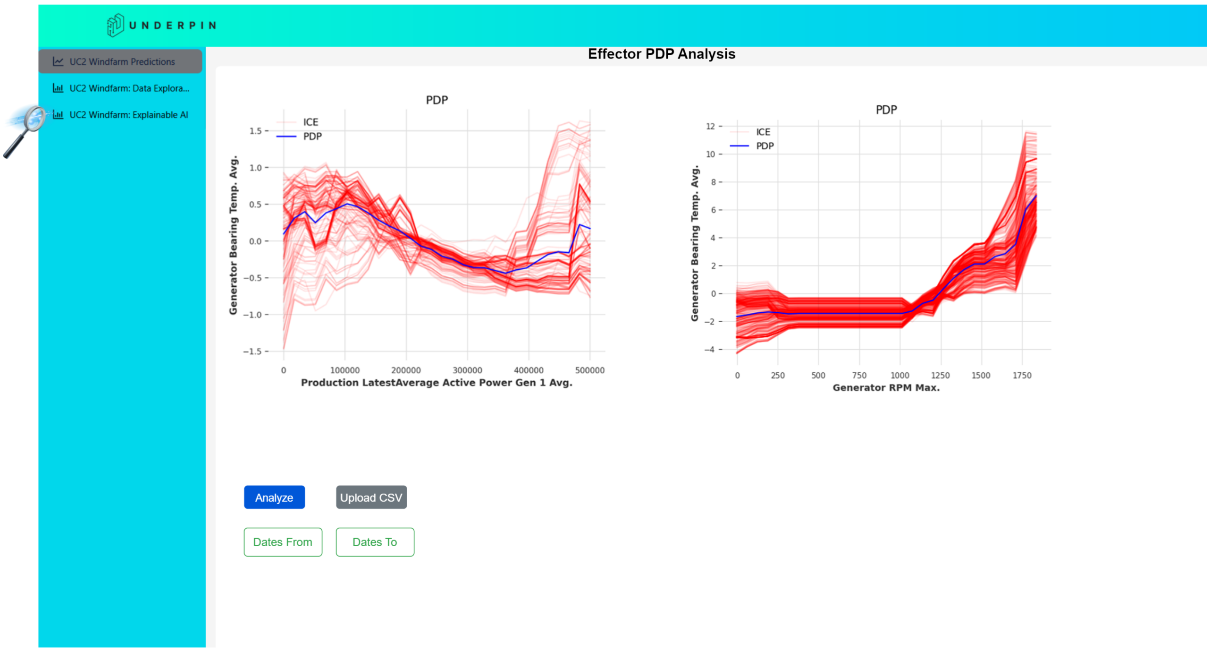

Figure 3: UNDERPIN Predictive Maintenance Dashboard

4. Results

4.1. PDP-ICE and SHAP Insights

In this illustration, we showcase the insights a user can gain when Effector is utilized to generate global effect plots. To obtain an explain of what happens in our models from input to output, we also use the SHAP Explainer in Darts. The SHAP method determines each feature's contribution to predictions, considering past lags of targets and covariates. SHAP values create a summary plot, ranking features by importance, aiding interpretation of the model's behavior. But SHAP has some significant limitations. SHAP shows ”what matters most” by ranking feature importance and contributes to the model's prediction, not to the real-world outcome. It cannot be used for causal analysis, and may highlight features that are important to the model but not to the underlying process. SHAP explanations may fail to capture how feature effects vary across subgroups or individual predictions, potentially masking important heterogeneity. On the contrary, PDP-ICE plots coming from Effector, show ”how it matters”, uncovering heterogeneous effects, detecting feature interactions, and shaping the actual relationships. Relying solely on SHAP can lead to misleading interpretations, especially in the presence of feature correlations or heterogeneous effects. Combining these tools provides a more robust and trustworthy approach to model interpretability in complex real-world scenarios.

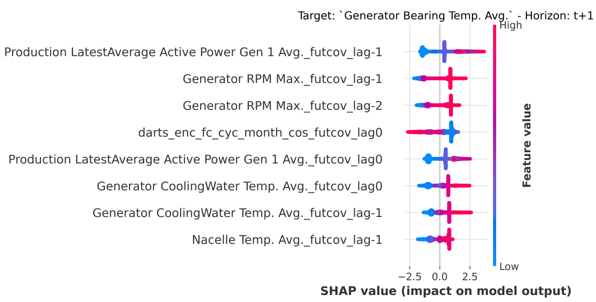

Figure 4: SHAP Summary Plot

The most influential features shown in the Figure 4 are the Production Latest Average Active Power Gen 1, Generator RPM Max and the Generator Cooling Water Temperature.

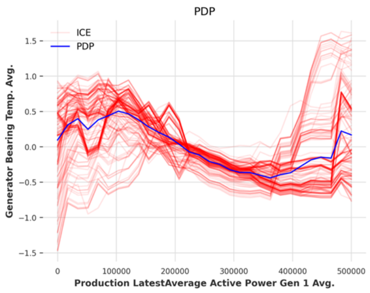

Figure 5: Active Power Generation (Watts)

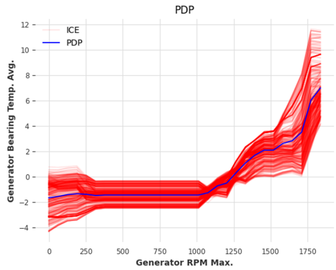

Figure 6: Generator RPM (revolutions/min)

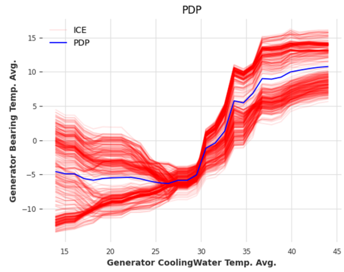

Figure 7: Cooling Water Temperature (°C)

4.2. Production Latest Average Active Power Gen 1 Analysis

The PDP-ICE analysis reveals critical insights for the Production LatestAverage Active Power Gen 1, identified as the top-ranked feature in SHAP importance analysis. Figure 5 demonstrates a complex multi-phase relationship with bearing temperature, exhibiting an initial rise during low power generation (0 - 100,000 units), followed by a decline to optimal efficiency zones (100,000 - 300,000 units), and a temperature increase at maximum power output (400,000+ units). The model identifies an optimal power generation range between 200,000 and 350,000 units where bearing temperatures are minimized and individual turbine variation is at its lowest, representing an operational sweet spot for mechanical efficiency.

Opposite to the general assumption of ”more power equals more heat”, the relationship reveals complex efficiency zones where mechanical systems exhibit non-linear thermal behavior, challenging conventional understanding of turbine operation. At maximum power output, individual turbines demonstrate extreme vulnerability with highly heterogeneous responses—some maintaining cool bearings while others experience significant heating, as evidenced by the substantial ICE plot divergence.

4.3. Generator RPM Max Analysis

The second most influential feature, Generator RPM Max, as ranked from Figure 4, shows in Figure 6 a clear threshold relationship where bearing temperatures remain stable below 1000 RPM but increase exponentially beyond this critical mechanical stress point, reaching 7-8°C at maximum RPM (2000). The ICE analysis demonstrates high individual variability at elevated RPM ranges, suggesting that mechanical stress tolerance varies dramatically between turbines, with implications for personalized operational limits.

4.4. Generator Cooling Water Temperature Analysis

The third most important feature, the Generator Cooling Water Temperature, exhibits a distinctive bimodal relationship. Figure 7 validates the mixed positive and negative SHAP values observed, where cooling water temperatures below 30°C effectively reduce bearing temperatures (reaching -6°C), while temperatures above this critical thermal threshold result in system overload conditions with bearing temperatures rising sharply to 10-15°C. This relationship reveals a critical cooling system effectiveness boundary at 30°C. Beyond that temperature, individual turbine cooling systems demonstrate extreme heterogeneity in performance—some maintain partial cooling effectiveness while others experience complete thermal management failure.

4.5. Analysis Conclusion

All three features exhibit consistent patterns where normal operating ranges show tight ICE clustering around the PDP (indicating predictable average behavior), while extreme operating conditions reveal substantial individual turbine heterogeneity, highlighting the necessity for turbine-specific monitoring strategies rather than fleet-wide averages. The temporal lag dependencies identified through SHAP analysis further emphasize that current bearing temperatures are strongly influenced by previous time-step power generation, RPM, and cooling system conditions, enabling proactive predictive maintenance interventions before critical thresholds are exceeded.

The Effector analysis revealed several key insights:

- Temperature-based features showed the strongest predictive power for anomaly detection
- Extreme heterogeneity identified under stress conditions
- Clear operating limits identified, showing consistent critical points at 30°C for temperature features and 1000 RPM for mechanical stress, indicating specific operating boundaries where performance changes dramatically and not gradually

4.6. Business Impact

The enhanced explainability enabled by Effector within the manufacturing data space environment allows maintenance teams to:

- Implement precision maintenance strategies based on individual turbine behavioral profiles revealed through ICE analysis.
- Optimize operational parameters using discovered efficiency zones (200,000 - 350,000 power range) and critical thresholds (30°C, 1000 RPM).
- Reduce false alarms by distinguishing between normal operational variations and genuine anomalies.
- Enable proactive interventions before critical thresholds are exceeded, leveraging temporal lag dependencies.
- Increase operational confidence through transparent AI explanations validated by domain experts.

The integration within the manufacturing Dataspace provides additional value to different users beyond maintenance teams. Dataspace users can benefit from a standardized access to explainable AI insights across organizational boundaries. This enables wind farm operators to benchmark turbine performance, equipment manufacturers to validate design assumptions using real operational data, and maintenance operators to make informed decisions based on transparent predictive maintenance schedules. The federated Dataspace architecture ensures that sensitive operational data remains secure while allowing collaborative insights that would be impossible in traditional siloed systems.

5. Conclusion and Future Work

This paper presented a successful real-world implementation of Effector XAI for industrial predictive maintenance in a manufacturing Dataspace environment. Our results demonstrate that regional explanations significantly improve the interpretability and actionability of machine learning predictions in complex industrial settings.

The integration of Effector within a manufacturing Dataspace highlights the potential of combining semantic technologies with explainable AI to create more transparent and trustworthy industrial AI systems. Future work will focus on extending this approach to multi-equipment scenarios and exploring the integration of temporal explanation methods for time series data.

The success of this implementation demonstrates that explainable AI tools like Effector provide essential operational insights into the model, making more clear the decision-making processes, which are of great relevance for the operators managing complex industrial systems. The additional insights into model behavior enable operators to make more informed maintenance decisions, ultimately bridging the gap between sophisticated AI predictions and actionable operational strategies.

Acknowledgments

This work is partially funded by the UNDERPIN project (Digital Europe Program, Grant Agreement no. 101123179). We also acknowledge the Effector development team for their support during the implementation process.

References

[1] International Data Spaces Association, International data spaces, 2025. URL: https://internationaldataspaces.org/, accessed: June 2025.

[2] C. Molnar, Interpretable Machine Learning, 2020. Https://christophm.github.io/interpretable-ml-book/.

[3] S. Lipovetsky, M. Conklin, Analysis of regression in game theory approach, Applied Stochastic Models in Business and Industry 17 (2001) 319–330.

[4] M. T. Ribeiro, S. Singh, C. Guestrin, "why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.

[5] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in Neural Information Processing Systems 30 (2017).

[6] J. H. Friedman, B. E. Popescu, Predictive learning via rule ensembles, The Annals of Applied Statistics 2 (2008) 916–954. URL: https://projecteuclid.org/journals/annals-of-applied-statistics/volume-2/issue-3/Predictive-learning-via-rule-ensembles/10.1214/07-AOAS148.full. doi:10.1214/07-AOAS148.

[7] A. Goldstein, A. Kapelner, J. Bleich, E. Pitkin, Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation (ice), Journal of Computational and Graphical Statistics 24 (2015) 44–65. URL: https://arxiv.org/abs/1309.6392. doi:10.1080/10618600.2014.907095.

[8] A. Author, Others, An explainable artificial intelligence model for predictive maintenance and spare parts optimization, Supply Chain Analytics 8 (2024). URL: https://avesis.marmara.edu.tr/yayin/782a6906-ff86-4fd2-82ed-94e1492cec2f/an-explainable-artificial-intelligence-model-for-predictive-maintenance-and-spare-parts-optimization.

[9] A. Ameli, et al., Xai for sensor fault detection in glass production, Sensors 22 (2022) 987.

[10] C. Senoner, T. H. Netland, S. Feuerriegel, Interpretable machine learning in manufacturing: Using shap to uncover process-quality relationships, Computers in Industry 140 (2022) 103674.

[11] L. Cummins, A. Sommers, S. Bakhtiari Ramezani, S. Mittal, J. Jabour, M. Seale, S. Rahimi, Explainable predictive maintenance: A survey of current methods, challenges and opportunities, arXiv preprint arXiv:2401.07871 (2024). URL: https://arxiv.org/abs/2401.07871.

[12] A. Author, Leveraging Explainable AI for Decision Support in Predictive Maintenance: a Framework for Operational Excellence, Ph.D. thesis, Politecnico di Milano, 2024. URL: https://www.politesi.polimi.it/handle/10589/227777.

[13] V. Gkolemis, P. Biecek, A. Kozak, P. Biecek, Effector: A python package for regional explanations, 2024. URL: https://arxiv.org/abs/2404.02629. arXiv:2404.02629.