INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 19, Issue: 3, Month: June, Year: 2024
Article Number: 6499, https://doi.org/10.15837/ijccc.2024.3.6499

CCC Publications

AGORA UNIVERSITY PRESS

Proposal of a Machine Learning Predictive Maintenance Solution Architecture

Aurelia Pătrașcu, Cristian Bucur, Ana Tănăsescu, Florentina Alina Toader

Aurelia Pătrașcu*
Economic Sciences Faculty
Petroleum-Gas University of Ploiesti, Romania
https://orcid.org/0000-0003-2890-5648
*Corresponding author: apatrascu@upg-ploiesti.ro

Cristian Bucur
Economic Sciences Faculty
Petroleum-Gas University of Ploiesti, Romania
https://orcid.org/0000-0002-7340-8870

Ana Tănăsescu
Economic Sciences Faculty
Petroleum-Gas University of Ploiesti, Romania
https://orcid.org/0000-0001-5965-3012

Florentina Alina Toader
Economic Sciences Faculty
Petroleum-Gas University of Ploiesti, Romania
https://orcid.org/0000-0002-2790-2579

Abstract

This article proposes an architecture model for a predictive maintenance solution that can be used to detect degradation of equipment in industrial units. This platform is compliant with Industry 4.0 standards and employs machine learning algorithms alongside with data analytics methodologies to model the kinetics of equipment degradation. This serves the overarching aim of Industry 4.0 by enabling real-time, data-driven decision making and complex asset management. To model the deterioration of the equipment, advanced data analysis techniques and machine learning were used, thus allowing for the early identification of imminent failures and reducing system downtime. The proposed solution was validated through numerous experiments and a comprehensive analysis of data. The results indicate not only enhanced operational reliability but also a reduction in environmental impact, thereby highlighting the value of intersecting Industry 4.0 and Sustainability paradigms in the field of industrial systems.

Keywords: predictive maintenance, machine learning techniques, platform architecture, data-driven approach, anomaly detection, Industry 4.0.

JEL Classification: C80, C63, C61

1 Introduction

The digitization of activities in manufacturing enterprises, as well as the increased use of smart devices and control software systems have led to the emergence of numerous data points within manufacturing execution systems (Morariu et al., 2020). The way enterprises can obtain value from processing large volumes of data and extract significant information provides a competitive advantage in terms of building control systems that ensures resource protection and production optimization.

Industry 4.0, a concept proposed at Hanovra in 2011, creates through smart enterprises "a world in which virtual and physical systems of manufacturing cooperate with each other in a flexible way" (Schwab, 2017). Industry 4.0 aims to build a digital representation of physical processes to understand much better how these processes unfold (Dalzochio et al., 2020; Panait et al., 2022). Thus, it connects machines, sensors and IT systems to enhance value across the supply chain (van Dinter, Tekinerdogan and Catal, 2022). These connected systems named cyber-physical systems can communicate with each other and transfer data by Internet information exchange protocols.

As production processes and equipment become more and more complex and intelligent, regulations have been elaborated that require the existence of safer and more reliable systems. The failure of a machine component can result in machine deterioration, environmental pollution and employees' injury. Therefore, maintenance is performed to prevent serious machine failure and to maintain its normal operational state (van Dinter, Tekinerdogan and Catal, 2022).

The predictive maintenance implementation can bring significant advantages to manufacturing enterprises. Furthermore, maintenance costs have an important share in total operating costs of such enterprises. Depending on the activity field, the share of these costs varies between 15 and 60% of the total costs (Mobley, 2002). However, companies do not accurately measure these costs and therefore it is necessary to use new technologies to monitor properly maintenance activity (Zonta et al., 2020).

Among the advantages of the predictive maintenance, the following are mentioned (Mobley, 2002; Dalzochio et al., 2020; Coandă, Avram and Constantin, 2020; Moldovan et al., 2022; Nunes, Santos and Rocha, 2023): decreasing the number of system failures caused by the incapacity to predict when its components will fail, increasing the productivity within the enterprise, reducing unplanned downtimes, efficient using of financial and human resources, decreasing costs to increase enterprises competitiveness and optimising maintenance interventions planning.

Different approaches that can be used to predict the failure of production equipment are described in the specialized literature. These are represented by physics-based approaches, knowledge-based approaches and data-driven approaches (Gao et al., 2015; Bektas, Marshall and Jones, 2020; Guo, Li and Li, 2020; Zhong, Han and Han, 2020; Zonta et al., 2020; Leohold, Engbers and Freitag, 2021; Soleimani, Campean and Neagu, 2021; Soualhi et al., 2022; van Dinter, Tekinerdogan and Catal, 2022; Nunes, Santos and Rocha, 2023).

Machine learning techniques can be used to predict and diagnose system failures by estimating the remaining lifetime of an equipment based on a large amount of data that trains the machine-learning algorithm (Dalzochio et al., 2020).

This article proposes a cloud-based architecture model for a predictive maintenance solution that can be used to prevent the degradation of pump seals in an industrial unit. Advanced data analysis and machine learning techniques have been leveraged in order to model the degradation of the pump seals, thus enabling early detection of impending failures and reducing system downtime. The proposed solution has been validated through many experiments and an extensive analysis of pump seal data. The study has made several key contributions to the domain of pump system analytics related to: *advanced data visualization, time-series analysis, feature engineering, anomaly detection and predictive modelling framework.*

2 Literature Review

The concept of "maintenance" is defined as an action performed to maintain a thing in its current state and to protect it from failures and deteriorations (Shetty, 2018).

According to the EN 13306:2017 standard, maintenance is "the combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function" (European Committee for Standardization, 2017).

Maintenance strategies have been modified over time based on the concerns and efforts of researchers, experts, engineers and technicians (Nunes, Santos and Rocha, 2023). Thus, these strategies have evolved (Shetty, 2018) from a reactive approach ("run to failure") to a predictive approach ("repair at the right time").

In the specialized literature, different types of maintenance strategies are presented (Shetty, 2018; Carvalho et al., 2019; Coandă, Avram and Constantin, 2020; Errandonea, Beltrán and Arrizabalaga, 2020; van Dinter, Tekinerdogan and Catal, 2022; Nunes, Santos and Rocha, 2023). Effective maintenance strategies allow: reducing unplanned production downtime, decreasing costs and increasing the useful lifetime of equipment used in industry (Nunes, Santos and Rocha, 2023). According to the classification proposed in the literature (Errandonea, Beltrán and Arrizabalaga, 2020; van Dinter, Tekinerdogan and Catal, 2022; Nunes, Santos and Rocha, 2023), maintenance strategies are classified as follows: reactive (corrective) maintenance, preventive maintenance, condition-based maintenance, predictive maintenance and prescriptive maintenance.

The study conducted in 2021 by IoT Analytics showed that predictive maintenance solutions were nearly non-existent in the market in 2015 (Brügge, 2021), predictive maintenance evolving from an approach based on statically monitoring of equipment conditions to a viable solution with a favourable return on investment.

The increased interest of researchers in the predictive maintenance field it is demonstrated in the specialized literature both by articles that perform a systematic literature review of the studies that treat this concept in the context of Industry 4.0 (Zonta et al., 2020; Sahli, Evans and Manohar, 2021) and the articles that use the same technique of systematic literature review to identify machine learning methods and techniques used in predictive maintenance (Carvalho et al., 2019; Dalzochio et al., 2020; Jain et al., 2022).

Predictive maintenance uses historical data, domain knowledge and models. This type of maintenance can predict behaviour models, trends and correlations through statistical models and machine learning methods, in order to identify equipment failures in advance, thereby improving decisions regarding maintenance activity and reducing downtime (Zonta et al., 2020). The concept of Internet of Things (IoT) plays an important role in the predictive maintenance process (Parpala and Iacob, 2017; Jiang, Wang and Jin, 2023; Simo et al., 2023). Industrial IoT (Boyes et al., 2018), a specialization of this concept, uses IoT technologies, machine learning methods and Big Data in the industry (Sezer et al., 2018).

Predictive maintenance is a broad research theme that uses knowledge from various fields of activity. To design an effective predictive maintenance system (Nunes, Santos and Rocha, 2023), data must be carefully processed to handle common industry situations (missing data, noise, erroneous values etc.).

Predictive maintenance is considered a component of the Prognostics and Health Management (PHM) system, along with Equipment Health Monitoring - EHM and Mean-Time-To-Repair – MTTR (Iskandar et al., 2015). Additionally, predicting the remaining useful life of equipment, that is a part of PHM (Zonta et al., 2020), is one of the main features of predictive maintenance (Nunes, Santos and Rocha, 2023). Some authors (Nunes, Santos and Rocha, 2023) connect this type of maintenance with the five features of Big Data. These features are known as the 5 V's of Big Data: volume, variety, velocity, veracity and value (Bazzaz Abkenar et al., 2021).

As it has been mentioned in the introduction, the approaches that can be used to identify the degradation of equipment can be classified in: physics-based approaches, knowledge-based approaches and data-driven approaches (Gao et al., 2015; Bektas, Marshall and Jones, 2020; Guo, Li and Li, 2020; Zhong, Han and Han, 2020; Zonta et al., 2020; Leohold, Engbers and Freitag, 2021; Soleimani, Campean and Neagu, 2021; Soualhi et al., 2022; van Dinter, Tekinerdogan and Catal, 2022; Nunes, Santos and Rocha, 2023).

Physics-based approaches use analytical patterns of physical systems behaviour that are based on probability distributions and laws of physics (Wu et al., 2018). Because these models describe physical processes, they directly and indirectly influence the operational state of equipment (Nunes, Santos and Rocha, 2023). Physics-based approaches can be used when precise mathematical models are available for a production system, based on its physical features (Soleimani, Campean and Neagu, 2021). These models provide a better way to handle the bias in the measured data and can describe the behaviour of a system, in different operating conditions. Additionally, the model's operation is transparent and easily interpretable because the fundamental equations and principles are explicitly represented (Gao et al., 2015). They do not require large volumes of data, making them suitable for systems where knowledge acquisition is costly (Soleimani, Campean and Neagu, 2021). However, these approaches also have disadvantages (Gao et al., 2015; Soleimani, Campean and Neagu, 2021; Soualhi et al., 2022): require detailed and complete knowledge of the system's behaviour, are application-specific and therefore cannot adapt to the system's changes, are computationally expensive and involve determining the mathematical model of the system.

Knowledge-based approaches decrease the complexity of physics-based models (Zonta et al., 2020) and use expert knowledge obtained based on the experience accumulated over years. The failures of a system, the deterioration state of an equipment or the root cause that generates the deterioration can be identified based on this knowledge (Nunes, Santos and Rocha, 2023).The models built based on expert knowledge can be classified as: rule-based models, case-based models and fuzzy logic-based models (Montero Jimenez et al., 2020). Knowledge-based models where the knowledge is represented by production rules have the advantage that the defined production rules are easy to read and the results are easy to understand and interpret (Leohold, Engbers and Freitag, 2021). They also involve less programming and data training, do not require the existence of models and can use incomplete and imprecise information (Soleimani, Campean and Neagu, 2021). One of the disadvantages of this type of model is that systems built using this approach are at risk of combinatorial explosions as the complexity of the systems and the volume of data increase. At the same time, these models rely on expert knowledge that is difficult for experts that build knowledge-based systems to understand and they do not learn from their own experience (especially rule-based systems). Another drawback of these approaches refers to the difficulty in acquiring knowledge from sources, as well as transforming it into the specific format of the system (Turban et al., 2004; Tănăsescu, 2016).

Data-driven approaches process data collected from sensors to extract significant knowledge from it (Nunes, Santos and Rocha, 2023). Data-driven approaches used to predict the deterioration of equipment are classified, according to the specialized literature (Nunes, Santos and Rocha, 2023), in: statistical methods and machine-learning methods (Bucur, 2019).

Data-driven approaches used to predict the equipment degradation have the following advantages: do not require knowledge of analytical models of equipment degradation, can be quickly and easily implemented, do not imply mathematical modelling of physical processes and establish relationships that were not previously identified by the acquisition of a large volume of data (Guo, Li and Li, 2020; Soleimani, Campean and Neagu, 2021; Soualhi et al., 2022). However, the performance of these approaches is dependent on the quality and quantity of data, making them inefficient in data-scarce environments (Leohold, Engbers and Freitag, 2021). Additionally, the results can be counter-intuitive when physical knowledge is not available (Soleimani, Campean and Neagu, 2021). Another weakness of these methods is that they require significant computation resources for training data, which may not be feasible in all operational settings (Soualhi et al., 2022).

3 Theoretical Framework

In the current landscape of technological advancement and intricate market dynamics, industrial organizations face the imperative challenge of boosting operational efficiency while ensuring economic sustainability. The exigency for economical asset utilization, process optimization, and operational safety is more critical than ever. This study posits that the deployment of sophisticated data analytics and machine learning paradigms can substantially enhance conventional maintenance and operational strategies within the industry.

The objective is to design a versatile software framework that is applicable across a spectrum of industrial contexts, from resource extraction in the oil and gas sectors to intricate manufacturing processes. This approach synergizes the strategic utilization of client data—both retrospective and real-time—with the forefront of computational technologies to bolster decision-making efficacy and operational productivity.

In the centre of our proposal is the conviction that data analytics, when leveraged to extract actionable insights, can significantly refine and advance industrial practices. The fundamental assertion is that the fusion of artificial intelligence and machine learning into industrial operations can unlock unprecedented strides in efficiency and dependability.

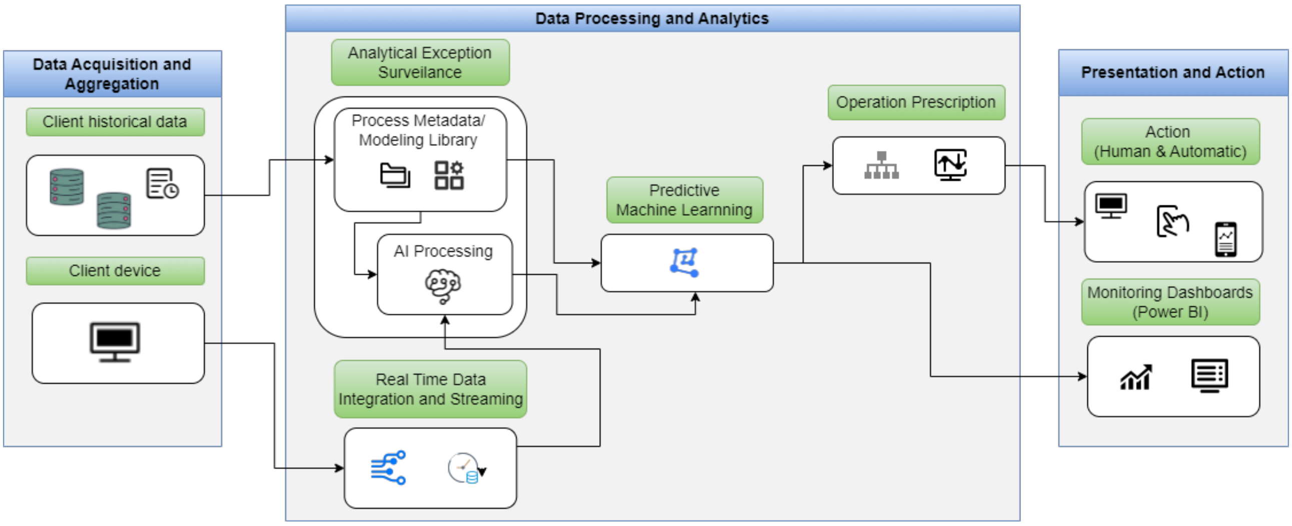


Figure 1: High-level Architecture of the Cloud-Enhanced Predictive Maintenance Solution.

Figure 1 illustrates the high-level architecture of our cloud-enabled solution, created to enable a comprehensive data lifecycle and informed decision-making in industrial settings. The architecture is segmented into three primary components: “Data Acquisition and Aggregation”, “Data Processing and Analytics” and “Presentation and Action”.

“Data Acquisition and Aggregation” module is pivotal for amassing and unifying diverse data streams. Historical data, indispensable for calibrating machine learning models and validating processes, consists of archival records of operational performance. Real-time data, harnessed from client devices like sensors, conveys the instantaneous operational state, enabling dynamic assessments and adjustments.

“Data Processing and Analytics” module serves as the analytical core and is where data is transformed into predictive insights:

- Real-Time Data Integration and Streaming: A cloud-based approach manages the ongoing surge of real-time data, ensuring high availability and scalability;
- Analytical Exception Surveillance: Harnessing cloud computing’s processing power, this component employs AI algorithms to identify and analyse anomalies and patterns that deviate from normative baselines;
- Predictive Machine Learning: Using cloud services, machine learning models are deployed that analyse historical and real-time data to predict trends and foresee potential operational interruptions;
- Operation Prescription: By leveraging the computational capabilities of the cloud, the system generates informed directives for operational management based on the analytics conducted.

In the “Presentation and Action” module, the insights are either visualized, using cloud-integrated dashboards like Power BI for decision-makers, or fed into automated systems that directly actuate operational adjustments, facilitating real-time responsive measures.

The overarching objective is to transition from reactive to proactive operational strategies in industrial environments. The underlying idea is that a data-driven, AI-enhanced approach can substantially reduce downtime, optimize resource utilization and lead to smarter, safer operational decisions.

This system is designed not just as a tool but also as a paradigm shift, encouraging a proactive, predictive approach to industrial maintenance and operations.

This research underscores that successful predictive maintenance transcends mere data analytics application. Essential is the integration of a platform that not only analyses but interprets data through the lens of domain expertise.

To practically execute predictive maintenance, a symbiosis between the technical acumen of engineers and a data analysis platform like the one presented in Figure 1 is needed. The domain expert or "knowledge engineer" are not confined only to technical knowledge and they apply a deep understanding of equipment operation to define optimized operational modes. In reality, equipment rarely runs strictly according to technical specifications due to various reasons, such as material variances, wear and tear, or dynamic operating conditions. Therefore, interpreting data through the lens of technical experience is indispensable for the continuous adaptation and optimization of maintenance processes.

This approach, enriched by the crucial contribution of domain expertise, forms the backbone of the proposed architecture, laying the foundation for a proactive maintenance strategy in the industrial environment. Thus, the described platform is not merely an analytical tool but an integrated ecosystem combining historical and real-time data with machine learning and technical knowledge, all via cloud infrastructure, to promote a new paradigm in industrial maintenance and equipment operation.

While the architecture follows a general framework for data acquisition, processing and visualization common in (big) data analytics, the novel contribution of our work lies in the specific integration of these stages and their application to industrial maintenance. Our system uniquely combines these processes into an agile, adaptive framework that leverages machine learning models, tailor-made for the context of industrial operations and maintenance, to foster a proactive rather than reactive strategy. The originality in the context of the provided platform lies within its integrated approach to predictive maintenance using real-time and historical data within a machine learning framework. The distinctiveness of this architecture is characterized by its ability to synthesize data from varied sources, apply advanced AI processing for anomaly detection, employ sophisticated predictive algorithms and effectively present actionable insights, all within a cohesive system.

4 Research Methodology

Based on the theoretical framework previously presented a test case has been performed. The case under scrutiny involves a Pump PM1A, component of an industrial complex unit. The pump serves the critical function of controlling unit temperature through injecting liquid. A recurring problem has been identified with respect to seal leakages, necessitating frequent and economically burdensome maintenance interventions, thereby influencing the mean time between failures (MTBF).

4.1 Data presentation and preprocessing

In this study, our data corpus was sourced from an extensive sensor network within the industrial unit, covering a period of four years. This network included approximately 30 sensors monitoring circular flow mechanisms and additional 40 sensors directly associated with the pump apparatus. The dataset with data collected at five-minute intervals provided a detailed temporal resolution that was important for our in-depth analysis. Data were extracted from the industrial unit's operational database, a repository of both historical and real-time operational data. The database provided comprehensive insight into the operational dynamics over an extended period.

The primary objective was to conduct a thorough examination of this high-dimensional dataset to identify key features or patterns that could inform proactive maintenance strategies. This involved detecting indicators of potential seal leakages and understanding the interdependencies within the sensor network that might influence the pump's operational health.

A critical aspect of our analysis focused on enhancing the Mean Time Between Failures (MTBF) Key Performance Indicator (KPI). By improving MTBF, we aimed to extend the operational lifespan of the pump seals. Notably, enhancing the MTBF of one component can lead to reduced stress and wear on other components, thereby enhancing the system's overall efficiency and lifespan.

By preprocessing and analysing this data, our study not only addressed the immediate challenge of pump seal failures, but also contributed to a broader understanding of system-wide operational efficiency, emphasizing the interconnection of individual component health and overall system performance.

In the process, data from multi-source heterogeneous environments is seamlessly integrated. The challenge of dealing with this heterogeneity, which arises from different variable formats, scales and temporal-spatial dimensions, is addressed through a meticulous preprocessing methodology. The initial data is collected from real-time sensors and features are extracted from Enterprise Resource Planning (ERP) systems. To ensure uniformity across the data, middleware solutions are employed to convert incoming data into a standardised format. As part of feature engineering, the data undergoes normalisation, logarithmic or exponential transformations and dimensionality reduction using Principal Component Analysis (Jolliffe, 2002). Data cleaning involves managing missing values and outliers. Resampling techniques are employed for time-series data to achieve uniform intervals, while data partitioning ensures a balanced class representation in both training and testing datasets, achieved through stratification.

4.2 Applied methodology

In order to achieve the proposed objectives, the first analytical phase involved a root cause analysis (Barsalou, 2014; Okes, 2019) employing reverse data-driven methodologies. Various machine learning models, reinforced with feature engineering techniques, were applied to determine the underlying causes of the seal leakages. The causal inference derived from this analytical foray served as the empirical basis for subsequent preventive measures.

The Reverse Data-Driven Root Cause Analysis model emphasised herein comprises a series of analytical stages, targeted at mitigating seal failures in the PM1A pump. Reverse Data Driven Root Cause Analysis (RCA) phases are:

- Identifying driving parameters through correlation analysis - analytical iteration (figure 2);
- Formulating root cause hypotheses by merging engineering knowledge with collected data;
- Adjusting failure model and closing the correlation gaps through RCA insights - analytical iteration (figure 3);
- Finalising descriptive model for seal failure mode (figure 4).

Identifying driving parameters through correlation analysis (analytical iteration) employs correlation analyses to isolate the critical operational parameters that have a direct or indirect relationship with seal degradation. This phase incorporates statistical tests to assess the significance of each identified parameter and may involve advanced multivariate analyses to elucidate potential interdependencies.

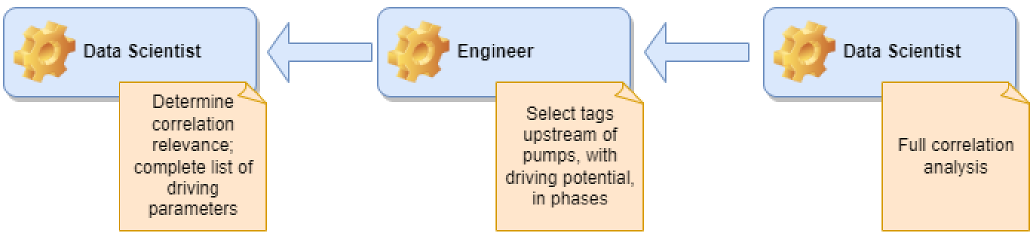


Figure 2: Identifying driving parameters through correlation analysis.

Formulating root cause hypotheses by combining engineering knowledge with collected data consists in merging the conventional engineering wisdom with the empirical findings from the data, thereby generating a set of root cause hypotheses. These hypotheses serve to identify the underlying mechanisms causing premature seal failures.

Adjusting failure model and closing correlation gaps through RCA Insights (analytical Iteration) uses the insights obtained from the root cause hypotheses to adjust and refine the seal life model. Specifically, this iteration aims to close the "correlation gaps", enabling a more accurate and nuanced understanding of seal degradation mechanisms.

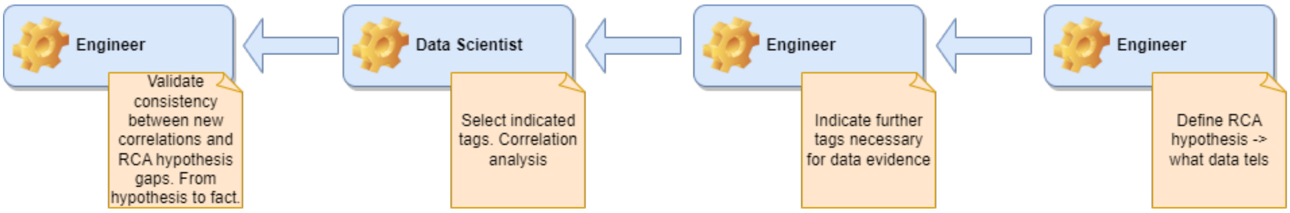


Figure 3: Adjusting failure model and closing the correlation gaps through RCA.

The terminal stage of RCA, *finalising descriptive model for seal failure mode*, elaborates the descriptive model for understanding seal failure modes. This model should ideally encapsulate all the salient driving parameters and their interrelationships into a coherent, predictive framework that allows for proactive maintenance decisions.

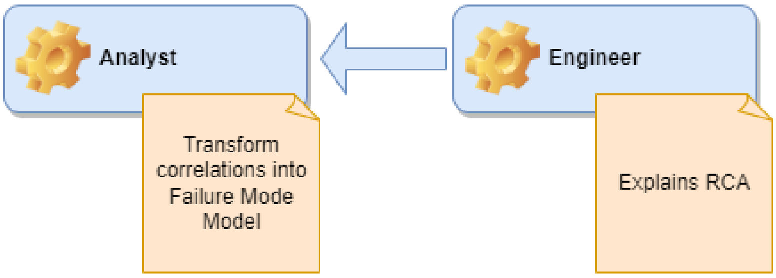


Figure 4: Finalising descriptive model for seal failure mode.

This systematic approach aims not only at discerning the root causes behind seal failures but also quantitative modelling of seal life.

The analysis proceeds with carrying out the first step of the RCA, when it is necessary to accomplish the task of determining the modes of operation (figure 5).

This task is instrumental in constructing an efficacious model for seal life expectancy in PM1A Pump. The methodology uses a fusion of empirical data and domain-specific knowledge, encompassing the following tasks:

- Determining Seal Life Cycles and Seal Break Events: This component employs a multi-dimensional analysis of historical data to delineate the life cycles of the pump seals. To achieve a robust interpretation, ERP reports are merged with domain-specific engineering knowledge. This synthesized dataset provides a nuanced picture of the temporal characteristics of seal degradation and failure, thus laying the groundwork for a comprehensive model.

- Identifying Modes of Operation of Interest: Subsequently, the historical data is scrutinized to classify various modes of pump operation. Specifically, the analysis distinguishes between effective operational states, stand-by states, as well as repair or maintenance phases. The identification of these modes is vital, as each mode imposes distinct stresses on the pump seals, thereby influencing their life expectancy.

In Figure 5, the pump operation modes are discerned through an examination of multi-dimensional sensor data streams. This data-driven analysis incorporates variations in temperatures, flow rates, speed and vibrations over time, all of which are critical indicators of the pump's operational state.

1. Temperature Data Analysis (Temperatures on the pump):

   Temperature sensors like TMP15PMA, TMP37PMA and TMP64PMA provide continuous readings. An 'Optimal TMP' threshold is established based on the pump's technical specifications

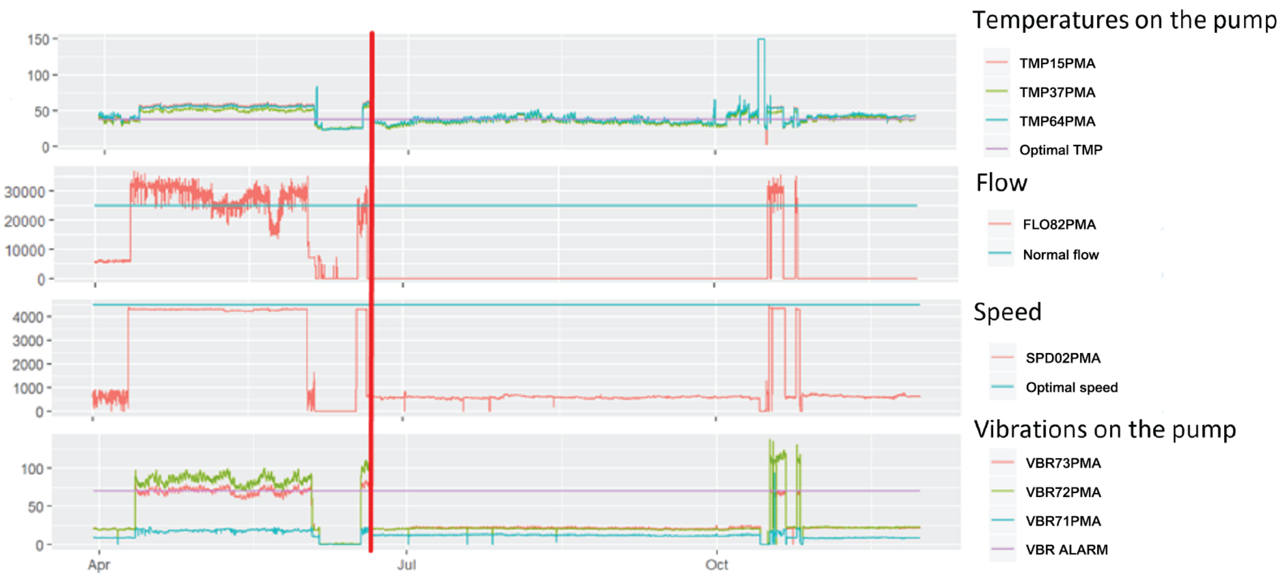


Figure 5: Determination of operation modes.

and historical performance data. Deviations from this optimal range can signify different operational states:

- Normal Operation: Temperatures consistently close to the 'Optimal TMP';
- Stand-by or Idle: Lower than normal temperature readings, suggesting reduced or no operational load;
- Maintenance or Repair: Significant temperature drops or unusual patterns may indicate the pump is offline for maintenance.

2. Flow Rate Examination (Flow):

   The flow sensor, FLO82PMA, informs on the volume of fluid passing through the pump. The 'Normal flow' line represents a benchmark for typical operations.

   - Full Operation: Readings that align with 'Normal flow'.
   - Reduced or Modified Operation: Readings below the normal flow could indicate a reduced demand or a pump operating in a non-standard mode, such as a low-efficiency state.

3. Speed Readings (Speed):

   Speed sensor, SPD02PMA, tracks the rotational velocity of the pump. An 'Optimal speed' line indicates the manufacturer's recommended operational speed.

   - Optimal Performance: Speed hovers around the 'Optimal speed'.
   - Reduced Speed Mode: Speed below the optimal can suggest energy-saving or limited operational mode.

4. Vibration Monitoring (Vibrations on the pump):

   Vibration sensors, VBR37PMA through VBR71PMA, monitor the pump's physical stability. A 'VBR ALARM' threshold alerts to potential mechanical issues.

   - Stable Operation: Vibration levels below the alarm threshold.
   - Potential Fault State: Spikes that reach the 'VBR ALARM' level may be symptomatic of mechanical faults, often necessitating a shift to maintenance mode.

Each of these data sets offers a composite view of the pump's operational state at any given moment. By establishing benchmarks and thresholds, the analytics system categorizes the operational modes. The analysis employs a combination of real-time and historical data to pinpoint patterns that correlate with specific operational states. By algorithmically clustering the data based on these thresholds and patterns, the system effectively segregates the operational timeline into discrete modes, such as 'Active Operation', 'Stand-by', 'Maintenance' or 'Fault State'. This stratification is crucial as it directly influences the predictive maintenance schedule and the life expectancy calculations for pump seals, informing maintenance decisions and optimization of operations. Further, the objective was to determine which upstream process parameters could act as precursors to anomalous behaviour in the pump system. This is a non-trivial task, implying a multivariate analytical approach that takes into consideration not only direct correlations but also time-lagged interactions and the weightage of each variable. This is executed in the initial phase of the RCA, specifically within Step 1.

4.3 Correlation model

Correlation, as defined within the realm of statistics, assesses the degree to which two variables fluctuate in tandem. Pearson's correlation coefficient (Boslaugh, 2012; Bucur et al., 2021) evaluates the strength and direction of the linear relationship between two quantitative variables. Pearson's correlation coefficient, symbolised as r, is calculated with the following equation:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{1}$$

where:
x and y represent the paired data sets being compared;
$\overline{x}$ and $\overline{y}$ indicate their respective means.

The resultant coefficient r ranges between -1 and 1, with the extremities indicating perfect linear relationships (negative or positive, respectively) and a value of 0 signifying no linear relationship.

For pump systems, understanding the Pearson correlation between variables (pump speed and temperature) and seal life can provide insights into optimal operating conditions. A strong negative or positive correlation might suggest that adjustments to these variables could prolong the seal life.

The analysis is based on an "all-with-all" correlation framework, examining pairwise correlations amongst the entire set of sensor metrics and operational variables. Special attention is accorded to the state when the pump exhibits what is termed "anomalous behaviour", which is precisely the condition wherein significant parameters deviate from their design-value ranges.

A detailed correlation analysis between vibration readings from the VBR70PMA sensor and densitometer readings from the DST series sensors within component unit D13 has been performed. This analysis was particularly insightful during instances of high vibrations, which were classified as abnormal states. Our experimental results are presented in next section: results and discussion.

5 Results and discussion

The quintessential aspect of this analytical iteration is the correlation of the contributors to this abnormal state, vis-à-vis upstream parameters that could act as potential early indicators or triggers. These correlations are scrutinised to be notably strong during these periods, thereby considering them as statistically significant.

For illustration, consider parameters related to pump vibrations and densitometer readings in component unit D13. During periods of anomalous behaviour typified by high vibration, certain parameters exhibit strong positive or negative correlations that are conspicuously absent or substantially weakened during periods of operational normality (figure 6).

For example, the VBR70PMA vibration sensor shows strong positive correlations with the other component unit parameters DST01PMA, DST01PMB and DST00PMC, registering correlation coefficients of 0.77, 0.91 and 0.93 respectively. This observation extends to other pairings as well, such as VBR71PMA with DST01PMA, DST01PMB and DST00PMC, having correlation coefficients of 0.81, 0.94 and 0.96, respectively.

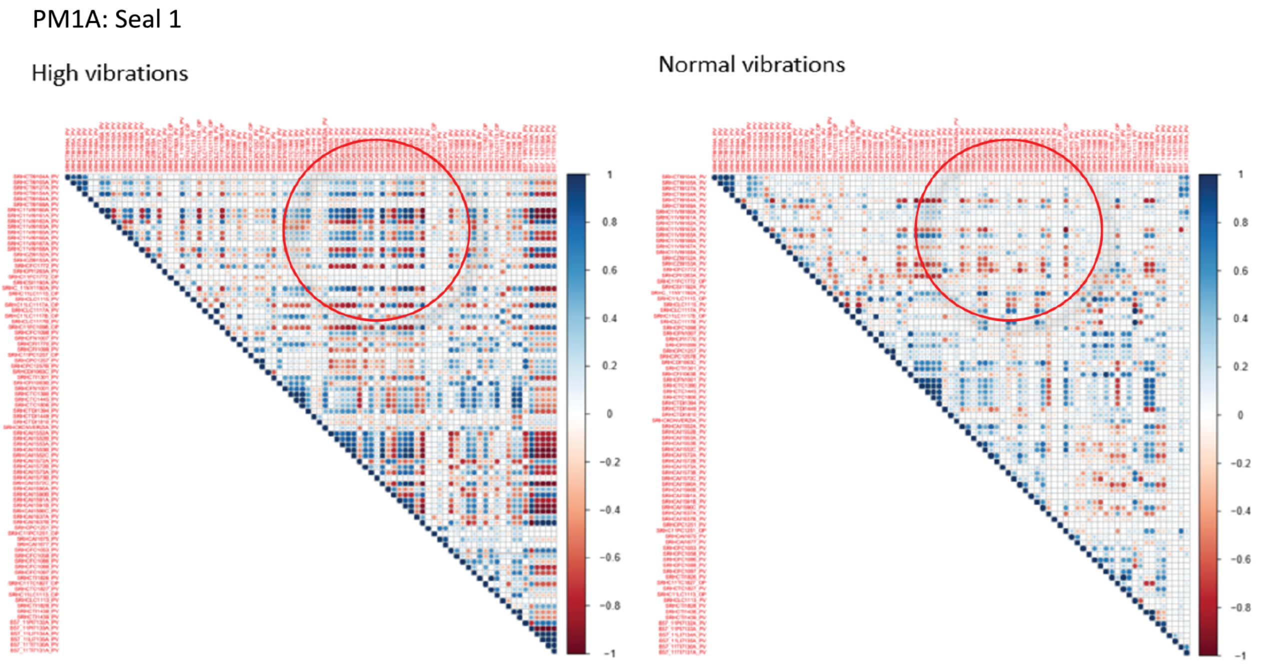


Figure 6: Correlations in normal and abnormal operation states.

In contrast, some parameters manifest a strong negative correlation during anomalous states, such as VBR70PMA and VBR71PMA with DST00PMB, registering coefficients of -0.78 and -0.80, respectively.

VBR70PMA – DST01PMA (0.77)
VBR70PMA – DST01PMB (0.91)
VBR70PMA – DST00PMC (0.93)
VBR71PMA – DST01PMA (0.81)
VBR71PMA – DST01PMB (0.94)
VBR71PMA – DST00PMC (0.96)
VBR70PMA – DST00PMB (-0.78)
VBR71PMA – DST00PMB (-0.80)

These observed correlations, which align with the anomalies circled in Figure 6, serve as indicators that may predict equipment malfunctions or imminent failures. By employing such a correlation model in practice, we can identify critical sensor pairings that exhibit strong correlations and could potentially serve as harbingers for predictive maintenance interventions.

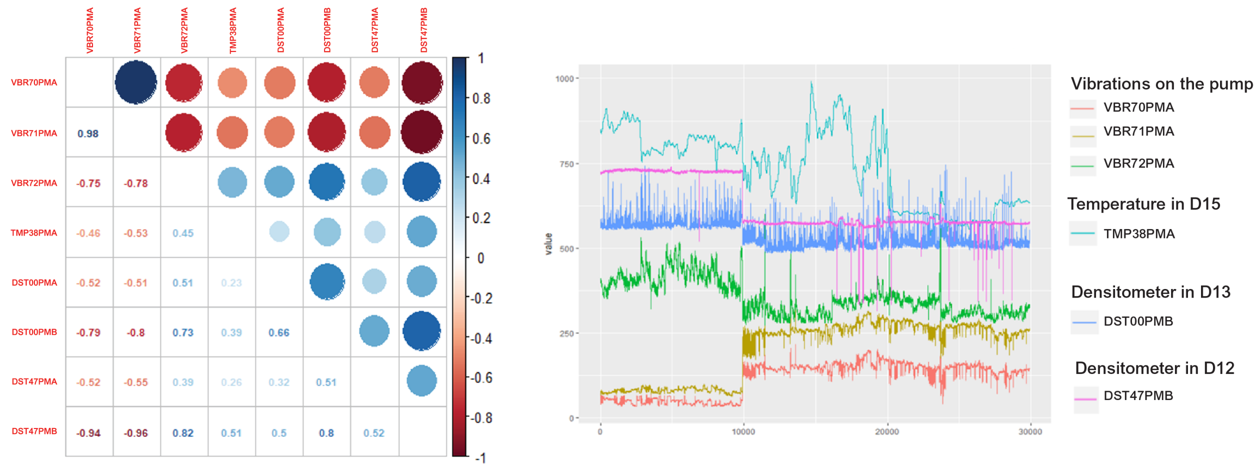
In the context of this study, there are emphasized two critical sets of variables: (a) holistic contributors and (b) driving parameters (figure 7).

On the one hand, the first set is represented by direct indicators of pump health and are closely linked to the Seal Life modelling. The second, on the other hand, is represented by upstream factors that, while not directly influencing the pump, have been identified through rigorous statistical analysis to significantly affect the holistic contributors to the failure model.

This methodology was applied to a real data set comprised of sensor readings from an active industrial pump system, captured and analysed over a significant period.

Holistic contributors, the first set of variables (figure 7a), include direct indicators such as vibration data from sensors VBR70PMA, VBR71PMA and VBR72PMA, and temperature readings from sensor TMP38PMA. These variables were directly correlated with the Seal Life Model, thus serving as immediate indicators for assessing pump health.

Driving parameters, the second set (figure 7b), encompass upstream factors, including density readings from densitometers DST00PMA and DST00PMB in components D13 and D22. While these do not directly impact pump performance, they were found, through meticulous statistical analysis, to significantly influence the holistic contributors. These findings are visually represented in the



(a) holistic contributors (b) driving parameters

Figure 7: Variables.

correlation matrix (figure 7a), which delineates the strength and direction of the relationships between each pair of variables.

This dual classification serves a vital function: the holistic contributors offer immediate markers for assessing pump states, while the upstream driving parameters act as precursors to changes in these states. Therefore, this architecture offers a nuanced analytical framework for predictive maintenance, determining an increase of MTBF and overall operational efficacy.

In situations characterised by emergent states, the analytical framework reveals intriguing patterns of interaction between high-density fine particles in component D13 and component D22 in industrial units (figure 8). Notably, during emergency states, the pump is relegated to stand-by status with diminished rotational speed, rendering vibration-based metrics non-informative. However, the correlation analysis between D13 and D22 densities, when cross-referenced with observations during regular operational modes, suggests a potential carryover of fine particles to the pump.

Such observations align with the dual-variable classification strategy. While vibrations serve as direct holistic contributors to the pump's seal life model during regular operations, the presence of high-density fine particles in D13 and D22 functions as a driving parameter that indirectly but significantly influences pump conditions during emergency states.

The data shows a clear link between increased density at the top of unit D13, which signals the presence of particles carry-over and increased vibrations and temperatures in unit D15 and the pump (figure 9). This correlation supports the idea that particles' carry-over could be affecting pump performance. By understanding these connections, better predictive models for maintenance can be achieved, improving both reliability and efficiency.

By corroborating these variables, it is obtained a multidimensional, high-resolution view into the systemic factors influencing pump performance. The elevated density at D13's apex serves as an upstream driving parameter that exerts a cascading influence on down-stream holistic contributors like vibrations and temperature metrics. It is noteworthy that the confluence of these variables, analysed in tandem, enhances the predictive potency of the machine learning models used for anomaly detection and predictive maintenance.

This demonstrates the utility of the cloud-based, data driven architecture in rendering actionable insights. The architecture allows for the synthesis of these seemingly disparate data points into a cohesive analytical model that accurately captures the complexities of the industrial system under consideration.

The data reveals a cyclical increase in temperature at unit D14, indicative of fouling phenomena. This temperature escalation is concomitant with elevated vibration and temperature metrics at the pump. Additionally, an increase in pump speed is observed, which can be interpreted as an operational manoeuvre to counterbalance the reduced cooling efficiency and deteriorating pump performance over time. These interconnected variables substantiate the hypothesis that component unit

Figure 8: High-density fine particles in D13 and D22.

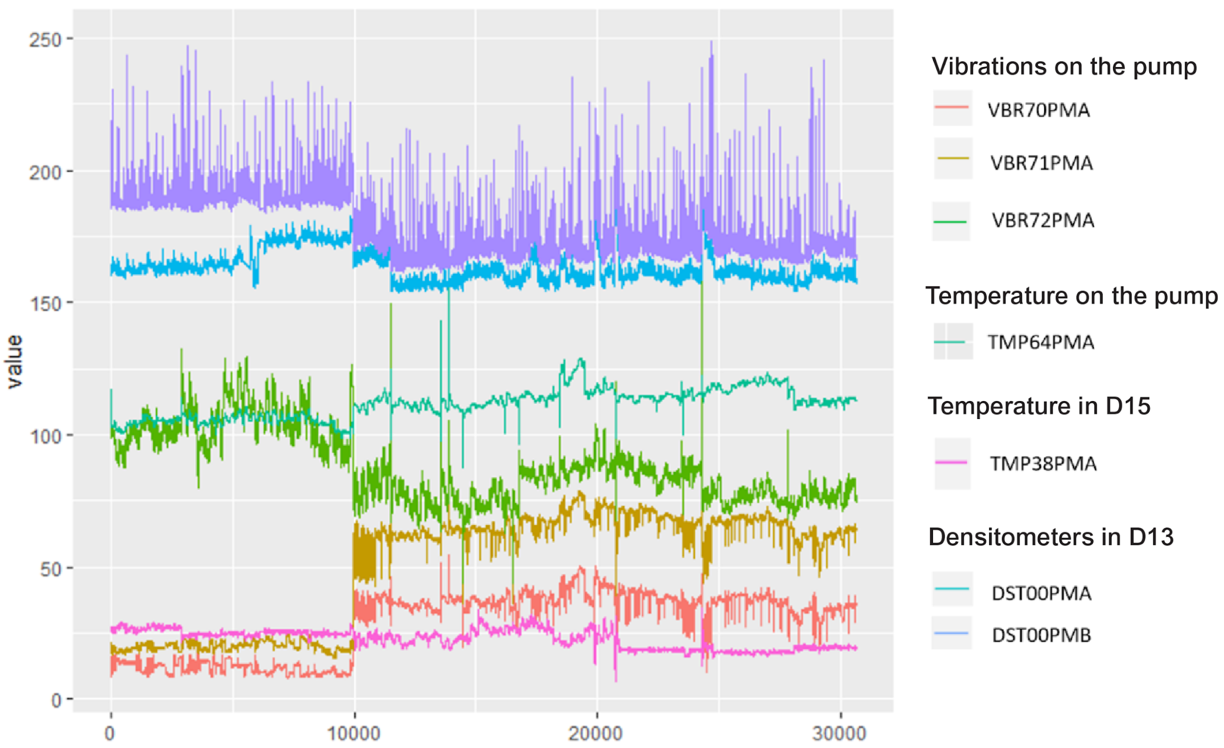


Figure 9: The increase of density, correlating with vibrations and temperature.

fouling adversely impacts pump operational integrity, thereby necessitating a more holistic approach in predictive maintenance algorithms to enhance asset longevity and operational efficacy.

In Figures 8 and 9, the high-density fine particles recorded in components D13 and D22, along with corresponding vibrations and temperature readings from the pump, serve as a case study highlighting a real-world application of our data-driven approach. During emergent states, when pumps enter a stand-by mode, we observe a reduction in vibration metrics, which typically signal operational issues. However, this does not render the system's data non-contributory. Instead, it directs our focus to other indicators, such as the density measurements in components D13 and D22, which maintain their diagnostic utility even when vibrational data becomes less informative.

The association between the heightened density of fine particles in these components and increased vibrations and temperature in the pump system is more than coincidental. It reveals an underlying mechanism affecting pump performance. The temporal patterns captured in the data reflect how particulate carry-over coincides with pronounced fluctuations in pump behaviour. This linkage underpins a predictive maintenance model that anticipates operational degradation, not solely based on individual sensor readings but through the synthesis of multi-sensor data, revealing a narrative about the pump system's health.

For instance, the data manifests cyclical temperature increases at unit D14, which are aligned with both increased vibrations at the pump and changes in operational speed. Such multi-variable correlation does not merely suggest, but rather evidences, the impact of fouling on pump efficiency and the reactive measures taken to mitigate this effect. This specific instance illustrates the application of data analytics to identify and respond to operational challenges in real-time, which is at the heart of our proposed solution.

By explaining these relationships and the analytical methods that uncover them, we provide a clear demonstration of how data-driven analysis directly informs maintenance strategies. The cloud-based architecture facilitates this by integrating, processing and analysing data from disparate sources to deliver actionable insights. Our approach is not only a generic endorsement of data analytics but also a practical deployment of these methods to address specific industrial challenges, which in this case, are centred on pump maintenance.

6 Conclusion

The presented architecture transcends the conventional boundaries of reactive maintenance strategies and embodies a comprehensive, cloud-based predictive maintenance framework tailored to the multifaceted demands of a diverse array of industrial stakeholders—ranging from floor operators and process engineers to reliability specialists and executive management. This novel approach harnesses the power of data-driven analytics and integrates it with the operational fabric of industrial maintenance.

What sets this architecture apart is its dynamic utilization of advanced machine learning algorithms. These algorithms delve deep into the operational data, not merely to predict potential system failures but to provide a granular understanding of asset health and process efficiency. This facilitates a strategic review in maintenance planning that smartly aligns with operational efficiency and maximizes asset lifespan. Real-time risk analytics—a cornerstone of the proposed system—offer continuous, across-the-board risk assessments, empowering stakeholders with the capability to dynamically prioritize risks and make informed decisions swiftly.

In terms of operational capabilities, the architecture presents multiple categories of benefits. Its algorithmic framework facilitates early detection of bottlenecks and equipment vulnerabilities. Additionally, one of its main advantages is the capacity to establish optimal operational limits under diverse conditions, thereby introducing a layer of operational flexibility unattainable until now. It also enables the predictive reshaping of maintenance schedules, a critical function for avoiding unscheduled downtimes. Furthermore, the architecture's analytics facility accelerates the diagnosis of process irregularities with unprecedented speed and accuracy. Finally, it incorporates a real-time criticality assessment module that contributes significantly to both immediate and long-term decision-making.

Implementing the architecture brings a suite of advanced functionalities into play:

- The ability to detect incipient issues and predict equipment vulnerabilities, leveraging machine learning and statistical insights;
- The capacity to dynamically calibrate operational procedures to maintain optimum performance under diverse conditions;
- The predictive scheduling of maintenance activities, minimizing the likelihood of unexpected system halts;
- The quick, precise diagnosis of operational irregularities, supported by a comprehensive analytical platform;
- An ongoing evaluation and management of the criticalities associated with process and equipment risks, reinforcing safety and operational guidelines.

The architecture heralds significant enhancements in several key areas:

- Enhanced operational efficacy through real-time monitoring and predictive analytics;
- Improved utilization of resources and assets, empowered by predictive insights;
- Elevated overall operational metrics, steered by data-driven decision-making processes;
- Strengthened safety measures and asset reliability, supported by immediate analytical feedback and early warning systems.

By integrating these functionalities, the architecture enables organisations to not only prevent and predict industrial process disruptions and facility outages, but also assess their criticality relative to achieving planned operational performance and risk mitigation. Furthermore, it establishes optimal preventive and contingency measures, completing the standard procedural framework in full compliance to existing regulatory norms. Therefore, it offers an integrated, data-driven paradigm for increasing the efficiency of operations and enhancing the value of industrial assets.

As we draw conclusions, it is imperative to reiterate the practical base of our architecture, which is validated using real operational data processed through a cloud-based solution. The architecture proposed has been brought to life by implementing it in an actual industrial setting, where the cloud infrastructure served as a backbone, supporting the intricate data workflows required for machine learning and predictive analytics.

Our affirmations are not merely theoretical postulations, but they are substantiated by empirical evidence derived from the deployment of our cloud-based framework. By capitalizing on the elasticity and scalability of cloud computing, we were able to manage large datasets, applying AI and machine learning algorithms to reveal patterns, trends and anomalies that form the core of our predictive maintenance strategy.

The results, illustrated through detailed visualizations, emphasize the efficacy of our approach. The real-world data, when fed through our cloud-implemented architecture, resulted in actionable insights that demonstrate the proactive capabilities of our system. These results reinforce our claims, underscoring the transformative potential of integrating advanced data analytics into industrial maintenance regimes. Through this fusion of cloud technology and data-driven methodologies, we have provided a solid foundation for our assertions, contributing a significant advancement in the field of industrial predictive maintenance.

Funding

This research was funded by a grant of the Petroleum-Gas University of Ploiesti, project number GO-GICS-11063/08.06.2023, within the Internal Grant for Scientific Research.

References

[1] Barsalou, M.A., 2014. Root Cause Analysis: A Step-By-Step Guide to Using the Right Tool at the Right Time. Boca Raton, Florida: CRC Press. https://doi.org/10.1201/b17834

[2] Bazzaz Abkenar, S., Haghi Kashani, M., Mahdipour, E. and Jameii, S.M., 2021. Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. Telematics and Informatics, 57, no. art.101517. https://doi.org/10.1016/j.tele.2020.101517

[3] Bektas, O., Marshall, J. and Jones, J.A., 2020. Comparison of Computational Prognostic Methods for Complex Systems Under Dynamic Regimes: A Review of Perspectives. Archives of Computational Methods in Engineering, 27, pp. 999–1011. https://doi.org/10.1007/s11831-019-09339-7

[4] Boslaugh, S., 2012. Statistics in a Nutshell. 2nd ed. Newton: O'Reilly Media, Incorporated.

[5] Boyes, H., Hallaq, B., Cunningham, J. and Watson, T., 2018. The industrial internet of things (IIoT): An analysis framework. Computers in Industry, 101, pp. 1-12. https://doi.org/10.1016/j.compind.2018.04.015

[6] Brügge, F., 2021. Predictive Maintenance Market: The Evolution from Niche Topic to High ROI Application, [online] Available at: <https://iot-analytics.com/predictive-maintenance-market-evolution-from-niche-topic-to-high-roi-application> [Accessed on 10 February 2023].

[7] Bucur, C., 2019. Using Machine Learning Techniques to Gain Insights from Enterprise Unstructured Data. Economic Insights – Trends and Challenges, 8(3), pp. 33-43.

[8] Bucur, C., Tudorică, B.G., Oprea, S.V., Nancu, D. and Duşmănescu, D.M., 2021. Insights into Energy Indicators Analytics Towards European Green Energy Transition Using Statistics and Self-Organizing Maps. IEEE Access, 9, pp. 64427-64444. https://doi.org/10.1109/ACCESS.2021.3075175

[9] Carvalho, T.P., Soares, F.A., Vita, R., Francisco, R.D.P., Basto, J.P. and Alcalá, S.G., 2019. A Systematic Literature Review Of Machine Learning Methods Applied To Predictive Maintenance. Computers & Industrial Engineering, 137, no. art.106024. https://doi.org/10.1016/j.cie.2019.106024

[10] Coandă, P., Avram, M. and Constantin, V., 2020. A state of the art of predictive maintenance techniques. In I. Doroftei, A. Popescu and C. Bujoreanu, eds. 2020. IOP Conference Series: Materials Science and Engineering. Bristol: IOP Publishing, vol. 997, art. no.012039. https://doi.org/10.1088/1757-899X/997/1/012039

[11] Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J. and Barbosa, J., 2020. Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. Computers in Industry, 123, art. no.103298. https://doi.org/10.1016/j.compind.2020.103298

[12] Errandonea, I., Beltrán, S. and Arrizabalaga, S., 2020. Digital Twin for maintenance: A literature review. Computers in Industry, 123, art. no.103316. https://doi.org/10.1016/j.compind.2020.103316

[13] European Committee for Standardization, 2017. EN 13306:2017 Maintenance - Maintenance terminology.

[14] Gao, R., Wang, L., Teti, R., Dornfeld, D., Kumara, S., Mori, M. and Helu, M., 2015. Cloud-enabled prognosis for manufacturing. CIRP Annals, 64(2), pp. 749–772. https://doi.org/10.1016/j.cirp.2015.05.011

[15] Guo, J., Li, Z. and Li, M., 2020. A Review on Prognostics Methods for Engineering Systems. IEEE Transactions on Reliability, 69(3), pp. 1110–1129. https://doi.org/10.1109/TR.2019.2957965

[16] Iskandar, J., Moyne, J., Subrahmanyam, K., Hawkins, P. and Armacost, M., 2015. Predictive maintenance in semiconductor manufacturing. In: Proceedings of the 26th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC). Saratoga Springs, NY, USA, 3–6 May 2015: IEEE. pp. 384-389. https://doi.org/10.1109/ASMC.2015.7164425

[17] Jain, M., Vasdev, D., Pal, K. and Sharma, V., 2022. Systematic literature review on predictive maintenance of vehicles and diagnosis of vehicle's health using machine learning techniques. Computational Intelligence, 38(6), pp. 1990-2008. https://doi.org/10.1111/coin.12553

[18] Jiang, Y., Wang, Z. and Jin, Z. (2023). Iot Data Processing and Scheduling Based on Deep Reinforcement Learning. International Journal of Computers Communications & Control, 18(6). https://doi.org/10.15837/ijccc.2023.6.5998

[19] Jolliffe, I., 2002. Principal component analysis. 2nd ed. New York, NY: Springer. https://doi.org/10.1007/b98835

[20] Leohold, S., Engbers, H. and Freitag, M., 2021. Prognostic Methods for Predictive Maintenance: A generalized Topology. IFAC-PapersOnLine, 54(1), pp. 629-634. https://doi.org/10.1016/j.ifacol.2021.08.073

[21] Mobley, R.K., 2002. An introduction to predictive maintenance. 2nd ed. Boston: Elsevier. https://doi.org/10.1016/B978-0-7506-7531-4.X5000-3

[22] Moldovan, O.G., Ghincu, R.V., Moldovan, A.O., Noje D. and Tarca, R.C. (2022). Fault Detection in Three-phase Induction Motor based on Data Acquisition and ANN based Data Processing. International Journal of Computers Communications & Control, 17(3). https://doi.org/10.15837/ijccc.2022.3.4788

[23] Montero Jimenez, J.J., Schwartz, S., Vingerhoeds, R., Grabot, B. and Salaün, M., 2020. Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. Journal of manufacturing systems, 56, pp. 539-557. https://doi.org/10.1016/j.jmsy.2020.07.008

[24] Morariu, C., Morariu, O., Răileanu, S. and Borangiu, T., 2020. Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems. Computers in Industry, 120, art. no.103244. https://doi.org/10.1016/j.compind.2020.103244

[25] Nunes, P., Santos, J. and Rocha, E., 2023. Challenges in predictive maintenance–A review. CIRP Journal of Manufacturing Science and Technology, 40, pp. 53-67. https://doi.org/10.1016/j.cirpj.2022.11.004

[26] Okes, D., 2019. Root cause analysis: The core of problem solving and corrective action. 2nd ed. Milwaukee, Wisconsin: ASQ Quality Press.

[27] Panait, M., Ionescu, R., Apostu, S.A. and Vasić, M., 2022. Innovation through Industry 4.0-Driving Economic Growth and Building Skills for Better Jobs. Economic Insights-Trends & Challenges, 11(2), pp. 109-117. https://doi.org/10.51865/EITC.2022.02.08

[28] Parpala, R.C. and Iacob, R., 2017. Application of IoT concept on predictive maintenance of industrial equipment. In I. Bondrea, C. Simion and M. Inţă, eds. 2017. MATEC Web of Conferences. EDP Sciences, vol. 121, art. no. 02008. https://doi.org/10.1051/matecconf/201712102008

[29] Sahli, A., Evans, R. and Manohar, A., 2021. Predictive maintenance in industry 4.0: Current themes. Procedia CIRP, 104, pp. 1948-1953. https://doi.org/10.1016/j.procir.2021.11.329

[30] Schwab, K., 2017. The fourth industrial revolution. New York: Penguin Random House.

[31] Sezer, E., Romero, D., Guedea, F., Macchi, M. and Emmanouilidis, C., 2018. An industry 4.0-enabled low cost predictive maintenance approach for SMEs. In: 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC). Stuttgart, Germany, 17–20 June 2018: IEEE. pp. 1–8. https://doi.org/10.1109/ICE.2018.8436307

[32] Simo, A., Dzitac, S., Duțu, A. and Pandelica, I. (2023). Smart Agriculture in the Digital Age: A Comprehensive IoT-Driven Greenhouse Monitoring System. International Journal of Computers Communications & Control, 18(6). https://doi.org/10.15837/ijccc.2023.6.6147

[33] Shetty, R.B., 2018. Predictive maintenance in the IoT era. In M.G. Pecht and M. Kang, eds. 2018. Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things. New York: John Wiley & Sons, Ltd. pp. 589-612. https://doi.org/10.1002/9781119515326.ch21

[34] Soleimani, M., Campean, F. and Neagu, D., 2021. Diagnostics and prognostics for complex systems: A review of methods and challenges. Quality and Reliability Engineering International, 37(8), pp. 3746–3778. https://doi.org/10.1002/qre.2947

[35] Soualhi, A., Lamraoui, M., Elyousfi, B. and Razik, H., 2022. PHM SURVEY: Implementation of prognostic methods for monitoring industrial systems. Energies, 15(19), art. no.6909. https://doi.org/10.3390/en15196909

[36] Tănăsescu, A., 2016. Sisteme inteligente în economie. București: Editura Universitară.

[37] Turban, E., Aronson, J.E. and Liang, T.P., 2004. Decision Support Systems and Intelligent Systems. 7th ed. Upper Saddle River, NJ: Prentice Hall.

[38] van Dinter, R., Tekinerdogan, B. and Catal, C., 2022. Predictive maintenance using digital twins: A systematic literature review. Information and Software Technology, 151, art. no.107008. https://doi.org/10.1016/j.infsof.2022.107008

[39] Wu, D., Jennings, C., Terpenny, J., Kumara, S. and Gao, R.X., 2018. Cloud-based parallel machine learning for tool wear prediction. Journal of Manufacturing Science and Engineering, 140(4), art. no.041005. https://doi.org/10.1115/1.4038002

[40] Zhong, K., Han, M. and Han, B., 2020. Data-driven based fault prognosis for industrial systems: A concise overview. IEEE/CAA Journal of Automatica Sinica, 7(2), pp. 330–345. https://doi.org/10.1109/JAS.2019.1911804

[41] Zonta, T., Da Costa, C.A., da Rosa Righi, R., de Lima, M.J., da Trindade, E.S. and Li, G.P., 2020. Predictive maintenance in the Industry 4.0: A systematic literature review. Computers & Industrial Engineering, 150, art. no.106889. https://doi.org/10.1016/j.cie.2020.106889

Cite this paper as:

Pătrașcu, A.; Bucur, C.; Tănăsescu, A.; Toader F.A. (2024). Proposal of a Machine Learning Predictive Maintenance Solution Architecture, *International Journal of Computers Communications & Control*, 19(3), 6499, 2024.
https://doi.org/10.15837/ijccc.2024.3.6499