

Article

Not peer-reviewed version

Analytical Framework for Reasoning-Based Downtime Forecasting

Neda Papić and [Mirjana Misita](#)*

Posted Date: 10 December 2025

doi: 10.20944/preprints202512.0919.v1

Keywords: reasoning models; downtime forecasting; machine systems



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Article

Analytical Framework for Reasoning-Based Downtime Forecasting

Neda Papić and Mirjana Misita *

University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia

* Correspondence: mmisita@mas.bg.ac.rs

Abstract

This study proposes an analytical framework that converts operational records into structured research hypotheses, offering a systematic approach to understanding failure mechanisms. The framework combines descriptive statistics, time-series analysis, linear regression, and machine-learning techniques to identify patterns, irregularities, and residual model behavior. Within the scope of this research, the framework was implemented as executable Python code and tested on a representative downtime category observed in a real industrial machine. The application demonstrated how analytical outputs can be translated into inductive, deductive, and abductive hypotheses that support deeper exploration of failure dynamics. The findings indicate the need for a comprehensive, multifaceted analytical approach to interpreting and forecasting machine downtimes, emphasizing that combining quantitative insights with the development of reasoning-based hypotheses increases the explainability and methodological rigor of the results.

Keywords: reasoning models; downtime forecasting; machine systems

1. Introduction

The ability to anticipate equipment failures is a fundamental prerequisite for ensuring operational continuity, enhancing safety, and reducing costs in modern industrial systems. With increasing automation and the growing availability of operational data, predictive maintenance has become essential for minimizing unplanned downtimes and improving asset reliability. Conventional data-driven approaches—based on statistical modeling and machine learning—can identify patterns in historical operational data. However, these methods often provide limited interpretative structure, especially in environments characterized by high variability, nonlinearity, and a lack of contextual information.

In practice, engineers and operators rely on different forms of reasoning to interpret early warning signals, identify potential failure mechanisms, and form expectations about future system behavior. Despite this, the systematic translation of analytical outputs into consistently formulated research hypotheses remains underdeveloped within existing predictive-maintenance methodologies. Prior literature has predominantly emphasized model accuracy and algorithmic complexity, while the methodological process of hypothesis formation—critical for understanding failure mechanisms—has received comparatively little attention.

This study addresses this gap by introducing an analytical framework designed to generate structured hypotheses from historical downtime data. By integrating descriptive statistics, time-series analysis, regression modeling, and machine-learning techniques, the framework enables the identification of operational patterns and provides a systematic basis for formulating inductive, deductive, and abductive assumptions regarding the causes and dynamics of failure events. The objective of this research is to support decision-makers in generating meaningful research hypotheses and to deepen the understanding of downtime phenomena in mechanical systems.

2. Background and Related Work

Monitoring and analysis of machine downtimes represent a critical component of industrial performance management, as unplanned interruptions directly increase production costs, reduce throughput, and compromise operational efficiency. While scheduled downtimes can be anticipated and planned around, unexpected failures introduce uncertainty into the manufacturing process and often lead to disproportionate losses in productivity. To manage these risks, organizations must systematically record and analyze historical downtime events, capturing not only their duration but also the operational conditions under which they occur. Such data enable the identification of patterns in the frequency and timing of failures, offering insight into how specific machine states, workloads, or environmental factors contribute to the emergence of unplanned stoppages. By understanding how often and under what circumstances particular failure modes arise, industrial practitioners can more accurately diagnose underlying issues, optimize maintenance strategies, and reduce the overall impact of unforeseen disruptions on production systems.

Despite the availability of large volumes of operational data, existing analytical approaches used in downtime assessment often struggle to capture the full complexity of machine behavior. Traditional statistical methods provide valuable summaries of failure distributions, yet they frequently overlook nonlinear interactions and contextual dependencies embedded in real operating environments. Similarly, machine-learning models can identify intricate patterns but typically function as black boxes, offering limited interpretability and insufficient guidance for understanding why failures occur. As a result, the process of translating raw analytical outputs into meaningful insights for decision-makers remains fragmented. This gap highlights the need for a structured methodology that not only quantifies downtime behavior but also supports the systematic formulation of well-founded hypotheses about the causes, mechanisms, and recurrence patterns of failures in mechanical systems.

Prior research has extensively examined the role of equipment condition monitoring and predictive maintenance in improving operational reliability. Foundational work in the field emphasizes that systematic tracking of machine condition and failure signatures is essential for reducing unplanned downtime and optimizing maintenance strategies [1]. Subsequent advancements in diagnostics and prognostics have focused on developing models capable of interpreting multi-sensor data, identifying degradation patterns, and supporting condition-based maintenance decisions across diverse mechanical systems [2]. With the rapid growth of data availability in industrial environments, machine-learning approaches have become increasingly prominent. Recent systematic reviews highlight the proliferation of algorithms designed to detect early signs of failure, classify fault types, and forecast future breakdowns, while noting that model performance strongly depends on data quality, feature selection, and contextual information [3]. Collectively, existing studies underscore the importance of integrating statistical, temporal, and data-driven techniques to gain meaningful insights into machine behavior—yet they provide limited methodological guidance on translating analytical outputs into structured research hypotheses. The present study addresses this gap by proposing an analytical framework that formalizes hypothesis generation from historical downtime data.

Windmark et al. [4] investigated downtime behavior in an industrial production system by analyzing the frequency, duration, and patterns of machine stoppages. Their study emphasized the importance of identifying dominant downtime categories and temporal structures in order to improve production planning, resource allocation, and maintenance strategies.

Inductive, deductive, and abductive reasoning represent three fundamental forms of inference that complement one another in both scientific research and everyday thought. Induction is based on deriving general conclusions from particular instances, with the strength of such conclusions being conditioned by factors such as the similarity and typicality of premises, as well as the diversity of examples [5,6]. Contemporary scholarship emphasizes that induction cannot be adequately explained solely through similarity-based models, but requires Bayesian and hybrid approaches that integrate formal precision with psychological plausibility [7,8]. Deduction, in contrast, proceeds from

general principles toward specific conclusions that are necessarily true if the premises are valid, which is why it is most often employed in hypothesis testing and the verification of theoretical assumptions [9]. Experimental studies demonstrate that deductive judgments are more strongly guided by the criterion of logical validity, whereas inductive judgments are more susceptible to heuristics and similarity [10]. This suggests that, while related, these two modes of reasoning rely on distinct cognitive processes. The third form, abduction, directs reasoning toward the best explanation and is especially salient when available data are incomplete. Abduction is employed across domains—ranging from causal explanation through formal logical systems [11], to interpretation in natural language processing [12], to linking empirical observations with theoretical models in logistics research [13]. Taken together, induction, deduction, and abduction constitute complementary epistemic pathways: induction enables generalization, deduction provides verification, and abduction opens space for novel explanations and hypotheses.

These three modes of reasoning find broad application across disciplines and are most often used in a complementary fashion. In technological and engineering contexts, inductive approaches have proven particularly valuable. Jurado et al. [14] developed fuzzy inductive models for fault prediction in smart grids, even under conditions of incomplete data, while Lu and Zhang [15] combined inductive learning with experimental design to enhance manufacturing planning. In education, Tóth and Pogatsnik [16] underscore that cultivating inductive reasoning among engineering students improves their capacity for solving complex problems, while similar principles underpin AI tools for production decision-making [17], where inductive models generate new knowledge. López Herrera [18] demonstrated that fuzzy inductive reasoning can successfully predict time series (e.g., water consumption and weather data), opening avenues for applications in smart sensors and control systems. Likewise, Filipič and Junkar [19] applied inductive machine learning in metal processing to support fluid classification and tool selection, thereby improving product quality and facilitating decision-making.

Deductive reasoning, by contrast, underpins robotics, where Nikitenko [20] showed that deductive modules generate new actions from general rules, while Ou et al. [21] demonstrated how deductive strategies maintain consistency in dialogue systems. In finance, Grek, Hartwig, and Dougherty [22] employed deductive models to test theoretical claims regarding optimal debt levels, while in procurement and marketing deduction validates decision rules identified through inductive observation [23]. Abduction, as highlighted by Kovács and Spens [13], plays a pivotal role in logistics research by connecting empirical observations with theoretical frameworks. In computer science, Hobbs et al. [12] showed that abduction aids in resolving ambiguity and generating interpretations in natural language understanding, while Shanahan [11] differentiated prediction as a deductive process from explanation as an abductive one. Similarly, in business and management contexts, Kaufmann, Wagner, and Carter [24] observed that abductive reasoning explains unexpected decision-making patterns, adding a complementary dimension to rational and intuitive approaches.

Contemporary research points to the need for deeper integration of inductive, deductive, and abductive reasoning within artificial intelligence, data analysis, and scientific modeling. In explainable AI (XAI), scholars emphasize moving beyond purely inductive methods toward systems that incorporate abductive mechanisms for generating human-understandable explanations [25]. Similarly, in time-series interpretation, researchers advocate abandoning the limitations of deductive classification systems in favor of abductive approaches that allow hypothesis generation and explanation even under conditions of incomplete or corrupted data [26].

In industrial applications, studies on predictive maintenance and production downtime forecasting underscore the role of advanced deep learning models (e.g., LSTM), but also the need for explainability and integration with human factors [27–29]. In this context, future research is expected to focus on balancing performance and interpretability to foster user trust in critical applications. At the theoretical level, there is considerable potential in developing multi-agent systems that integrate diverse forms of reasoning to produce more stable and interpretable decisions. Frameworks such as the Theorem-of-Thought model illustrate how inductive, deductive, and abductive approaches can

be orchestrated into formal reasoning graphs, promising more robust systems in language processing and logical inference [30].

Moreover, inductive methods are expected to gain further importance in big data research and in predicting relationships within complex structures such as knowledge graphs [31,32]. Finally, in education and the development of scientific reasoning, the role of abductive reasoning in modeling competencies has been recognized as a key direction for teaching students to connect theories and empirical data [33].

3. Methodological Framework

The proposed analytical framework, Figure 1, consists of six sequential steps designed to extract informational value from historical downtime data and translate it into structured research hypotheses. Each step contributes a distinct analytical perspective that jointly enables the systematic exploration of failure patterns.

The analysis begins with a descriptive statistical assessment of the dataset, including measures such as the mean, minimum, maximum, standard deviation, and coefficient of variation. This step characterizes the basic structure, dispersion, and distributional properties of downtime intervals and durations, providing an essential foundation for all sub-sequent analyses.

The second step examines temporal behavior in the occurrence and duration of downtimes. Techniques such as exponential smoothing are applied to identify trends, volatility, and recurring dynamics. Forecasting errors, expressed through RMSE, are used to evaluate the extent to which temporal patterns contribute to understanding failure phenomena.

The third step assesses whether a linear relationship exists between event sequence and downtime characteristics. Linear regression models are used to identify directional trends and to quantify explanatory power using performance metrics such as R^2 and RMSE.

To expand the analytical scope beyond the limitations of simple linear relationships, the framework incorporates machine-learning models such as Random Forest, Support Vector Regression, Decision Trees, and related algorithms. These models allow for the exploration of complex interactions, structured variability, and latent patterns within the data that may not be adequately captured through linear regression alone.

Insights obtained from the preceding analytical steps are synthesized to formulate inductive, deductive, and abductive hypotheses. Inductive hypotheses arise from observed statistical regularities; deductive hypotheses are derived from rule-based or threshold-based inferences; and abductive hypotheses emerge from anomalies, residual unexplained variance, or unexpected model behavior.

In the final step, the dataset is partitioned into training (70%) and test (30%) subsets to allow for empirical validation of all generated hypotheses. Each hypothesis is evaluated by examining how well the corresponding analytical or predictive model performs on previously unseen data. Performance metrics such as predictive accuracy and mean squared error are used to determine the extent to which the empirical evidence supports the proposed assumptions. The testing procedure results in a structured summary of hypothesis validity and model performance, providing a foundation for deeper investigation of downtime mechanisms and for guiding subsequent research and decision-making.

Based on the statistical and predictive analyses, a set of reasoning-based hypotheses is constructed, denoted by:

$$H = \{H_1, H_2, \dots, H_{10}\} \quad (1)$$

These hypotheses are organized according to their inferential structure::

1. Inductive ($H_1 - H_4$): based on observed statistical regularities and model outputs;
2. Deductive ($H_5 - H_7$): based on logical rules and operational thresholds;
3. Abductive ($H_8 - H_{10}$): aimed at explaining anomalies and model failures.

Each hypothesis $H_j \in H$ is evaluated on the test set using logical or numerical criteria. The evaluation metrics include: accuracy and MSE.

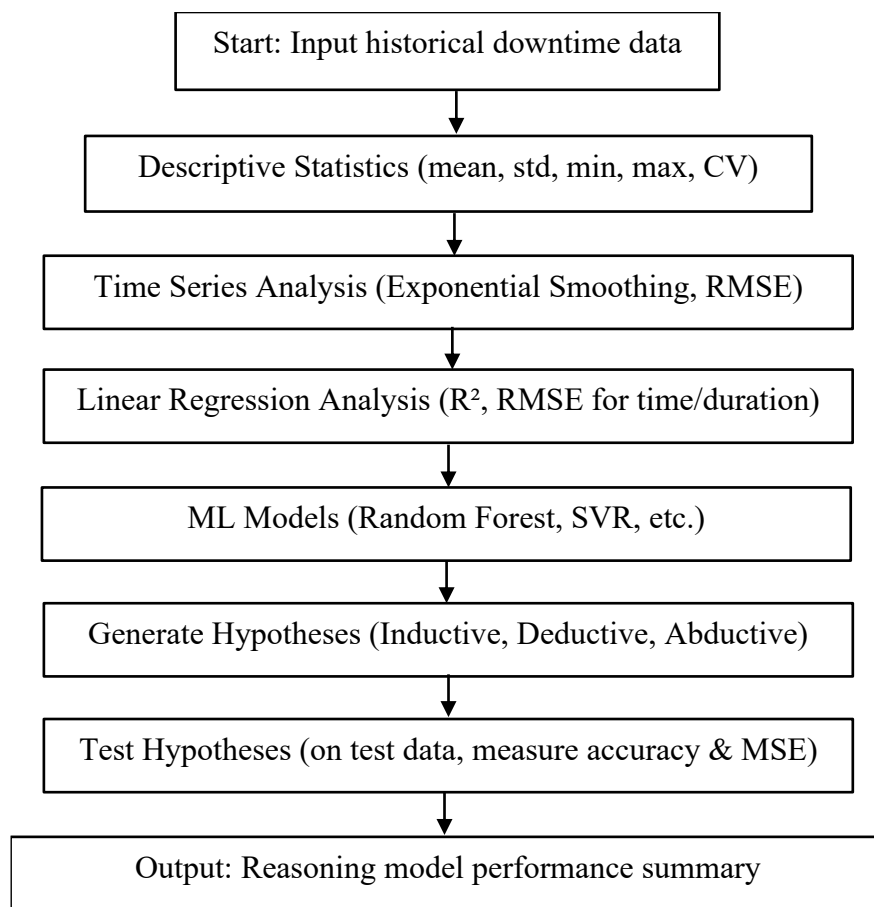


Figure 1. An analytical framework for reasoning-based downtime forecasting.

Accuracy, calculated as: Each hypothesis H_j is evaluated on the test dataset using either logical or numerical criteria. Quantitative evaluation metrics include accuracy

$$H_j = \frac{1}{N} \sum \mathbf{1}\{H_j(x_i) = t_i\} \quad (2)$$

MSE, if the hypothesis yields numerical predictions:

$$MSE(H_j) = \frac{1}{N} \sum (\hat{y}_i - y_i)^2 \quad (3)$$

These metrics quantify the empirical support for each hypothesis on unseen data. The integrated evaluation procedure ensures that hypotheses are assessed according to their evidential basis and their contribution to a deeper understanding of downtime mechanisms.

4. Results

In this experimental study, the operation of an industrial machine was monitored continuously over a six-month period, during which each downtime event and its duration were manually recorded. More than twenty distinct downtime categories were observed. For the purpose of validating the analytical framework proposed in this study, one category of unplanned downtime was selected to examine the occurrence patterns of a specific failure mode and to derive hypotheses related to its underlying mechanisms.

To support hypothesis development, the analysis was further focused on downtime events associated with tool wear, degradation, or breakage. Concentrating on this failure type enabled a

detailed investigation of a phenomenon that is highly relevant for as-sessing machine reliability and for informing maintenance planning.

Table 1 presents the basic descriptive statistics for the selected downtime category. The dataset includes the total number of downtime events, their duration, and the time between failures, as well as key statistical indicators such as the arithmetic mean, mini-mum and maximum values, standard deviation, and coefficient of variation. These measures provide a quantitative foundation for subsequent modeling and forecasting of machine downtime.

Table 1. Descriptive Statistics of Downtime Due to Tooth Wear.

| Metric | Time Between Failures (min) | Downtime Duration (min) |
|--------------------------|-----------------------------|-------------------------|
| Number of Events | 40 | 40 |
| Mean | 7,726.25 | 53.25 |
| Minimum | 0 | 10 |
| Maximum | 34,610 | 205 |
| Standard Deviation | 9,124.114 | 42.27 |
| Coefficient of Variation | 1.181 | 0.79 |

Time series analysis using the Exponential Smoothing model was conducted to identify recurring patterns in the timing and duration of unplanned downtimes. The goal was to establish a foundation for forecasting future events based on historical dynamics. Figure 2 illustrates the observed fluctuations in time between failures and corresponding downtime durations.

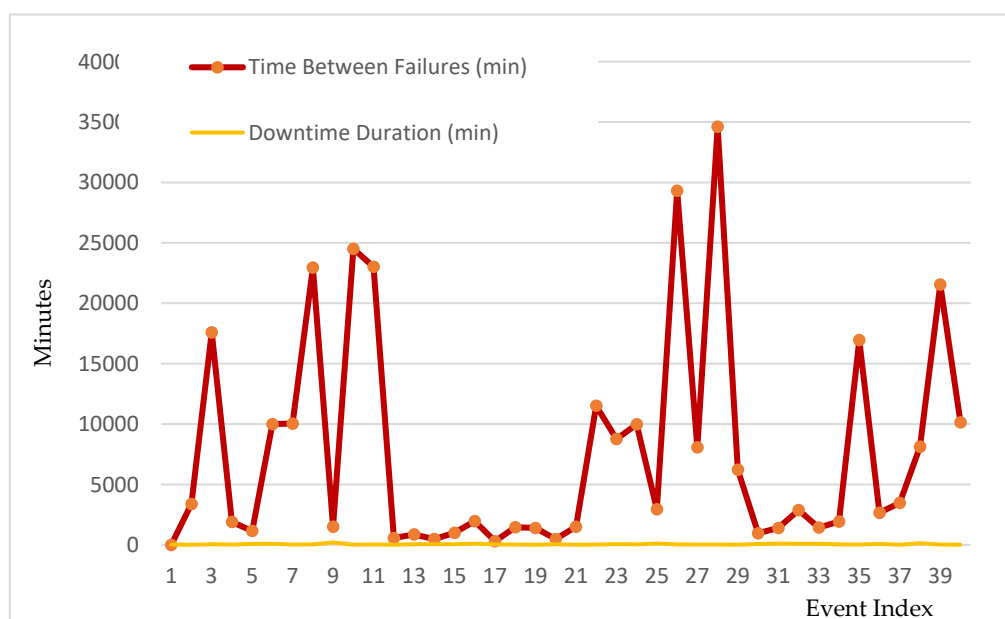


Figure 2. Downtime Forecasting Time Series of Time Between Failures and Downtime Duration.

4.1. Time Series Analysis Using Exponential Smoothing

The time series analysis using exponential smoothing produced a forecasted time until the next failure of 9,834 minutes, which corresponds to approximately 164 hours. The expected duration of the upcoming downtime was estimated at 54 minutes. Model accuracy was evaluated using root mean squared error (RMSE), yielding 9,653 minutes for time between failures and 45 minutes for downtime duration. These error values indicate substantial variability in the data, suggesting that the time series model struggles to reliably predict future events based solely on historical time patterns.

In the subsequent phase of the research, the linear regression method was applied to evaluate the potential for predicting unplanned downtime based on prior patterns. Figure 3 illustrates the results of the linear regression model applied to predict unplanned down-time due to tooth wear. The model uses a simple linear function to predict future values based on past observations.

4.2. Linear Regression Model

The linear regression model yielded a predicted time until the next failure of 8,812 minutes and an expected downtime duration of 57 minutes. The model's accuracy was evaluated using standard error metrics, resulting in a root mean squared error (RMSE) of 8,988.6 minutes for time between failures and 41.7 minutes for downtime duration. The coefficient of determination (R^2) for both predictions was approximately 0.0046, indicating a very low explanatory power of the linear trend in capturing the variability of the observed data.

4.3. ML Models

In the subsequent phase of the research, machine learning models were applied to the prediction of time between failures and downtime duration. Table 2 summarizes the performance of these models, while Figure 3 presents a comparative visualization of their predictive accuracy based on R^2 values.

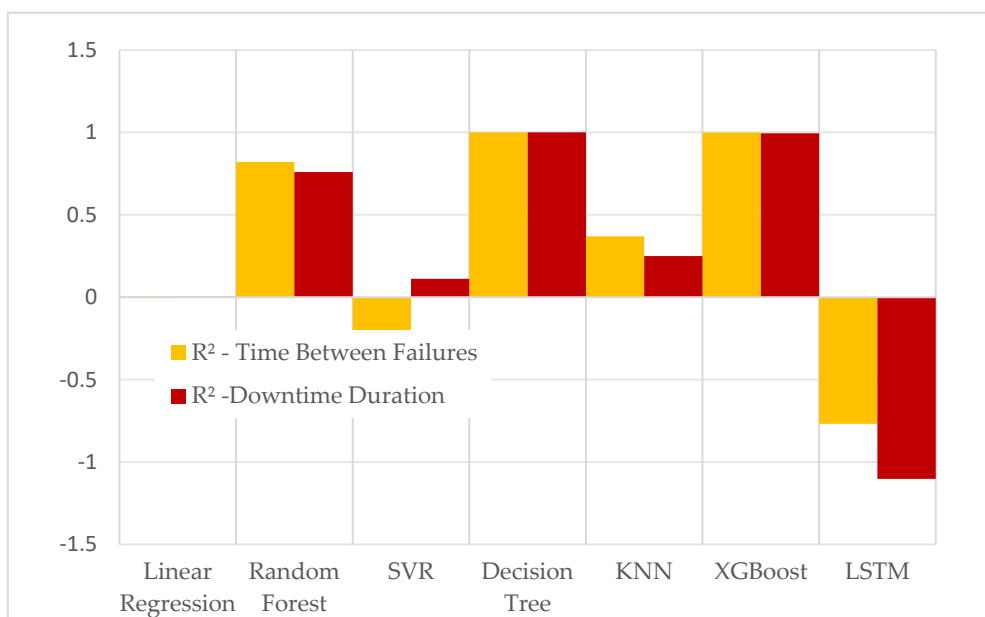


Figure 3. Comparative Performance of ML models.

Table 2. Machine Learning models Due to Tooth Wear.

| Model | MSE (Time) | R^2 (Time) | MSE (Duration) | R^2 (Duration) |
|-------------------|----------------|--------------|----------------|------------------|
| Linear Regression | 80794752.2221 | 0.0046 | 1736.7169 | 0.0030 |
| Random Forest | 14239393.6650 | 0.8246 | 418.4464 | 0.7598 |
| SVR | 102958977.0961 | -0.2685 | 1547.8136 | 0.1114 |
| Decision Tree | 0.0000 | 1.0000 | 0.0000 | 1.0000 |
| KNN | 51175693.6111 | 0.3695 | 1307.2917 | 0.2495 |
| XGBoost | 11828.3723 | 0.9999 | 9.2271 | 0.9947 |
| LSTM | 144441860.6277 | -0.7684 | 3712.5785 | -1.1027 |

The evaluation of machine learning models demonstrated highly uneven predictive performance across the different approaches. Linear Regression, SVR, KNN, and particularly LSTM yielded very low or even negative R^2 values, indicating that these models were largely unable to capture the variability in either time between failures or downtime duration. Although the Decision Tree achieved perfect R^2 and zero MSE on the training set, this outcome clearly reflects overfitting rather than genuine predictive ability.

In contrast, the ensemble-based models—most notably XGBoost and Random Forest—exhibited substantially stronger performance. XGBoost achieved extremely low MSE values ($\approx 11,828$ for time between failures and ≈ 9 for downtime) and near-perfect R^2 scores (above 0.99 for both outputs). Random Forest also captured a notable portion of variance, with R^2 values above 0.82 for time between failures and approximately 0.76–0.79 for downtime duration, depending on the training instance.

These results indicate that nonlinear ensemble methods are far better suited for modeling the irregular, noisy, and highly variable failure patterns present in the dataset, whereas simpler linear or distance-based methods fail to capture the underlying structure. Overall, the findings reinforce that the time-indexed sequence alone is insufficient for reliable prediction, suggesting that critical contextual and operational factors remain unobserved. This underscores the importance of integrating richer input variables and abductive reasoning within future predictive frameworks.

4.4. Reasoning-Based Models

Drawing upon the results obtained from descriptive statistical analysis, time series modeling, linear regression, and machine learning techniques, the dataset was systematically examined in terms of its variability, structural patterns, and underlying failure dynamics. These analytical procedures provided the empirical basis for formulating a set of reasoning-based hypotheses—inductive, deductive, and abductive—designed to capture different dimensions of the observed downtime phenomena. The resulting hypotheses, which reflect statistical regularities, operational rules, and plausible explanatory mechanisms, are summarized in Table 3 according to their corresponding reasoning type.

Table 3. Reasoning-based models.

| Type | Code | Hypothesis | Source / Derivation |
|-----------|------|--|--|
| Inductive | H1 | The observed average time between failures due to tooth wear is approximately 7,726 minutes. | Derived directly from descriptive statistics |
| Inductive | H2 | The average downtime duration due to tooth wear is 53 minutes. | Derived from descriptive statistics |
| Inductive | H3 | Most ML models (Linear Regression, SVR, LSTM) demonstrate low or negative R^2 , while KNN shows only modest explanatory power. | Derived from ML model performance |
| Inductive | H4 | XGBoost achieves the lowest MSE and highest R^2 among all valid (non-overfitting) models. | Derived from ML model performance |

| | | | |
|-----------|-----|---|---|
| Deductive | H5 | If the operational interval approaches or exceeds ~9,000 minutes, a failure event becomes likely. | Based on exponential smoothing forecast ($\approx 9,834$ min) and linear regression forecast ($\approx 8,812$ min). |
| Deductive | H6 | If a downtime exceeds 100 minutes, it represents a significant deviation from expected duration. | Derived from descriptive mean of 53 min and forecasted durations (54–57 min). |
| Deductive | H7 | The sequence of failures is not informative for prediction. | Based on near-zero R^2 in time-series and linear regression models ($R^2 \approx 0.0046$). |
| Abductive | H8 | Unplanned downtime due to tooth wear is influenced by operational or environmental factors not captured in the dataset. | Based on poor performance of most models and time-series limitations (high RMSE, low R^2). |
| Abductive | H9 | Progressive micro-damage accumulation contributes to eventual failure, beyond what temporal patterns can explain. | Derived from model inability to infer degradation from time alone (time-series RMSE $> 9,500$). |
| Abductive | H10 | Tool wear likely follows non-linear or cyclic degradation, not strictly time-based. | Based on strong performance of nonlinear models (XGBoost/RF) vs. failure of linear/time-series models. |

4.5. Hypothesis Testing

The results of the hypothesis testing are systematically presented in Table 4, which provides an overview of the validity of each hypothesis together with the corresponding evaluation methods and interpretations.

Table 4. Hypothesis Testing.

| Code | Hypothesis | Test Method | Result | Interpretation |
|------|---|---------------------------------|----------------------------|----------------|
| H1 | Average time between failures $\approx 7,726$ min | One-sample t-test | $t=0.000$, $p=1.000$ | SUPPORTED |
| H2 | Average downtime ≈ 53 min | One-sample t-test | $t=0.000$, $p=1.000$ | SUPPORTED |
| H3 | Most ML models have $R^2 < 0.05$ | Descriptive comparison of R^2 | 3/7 models below threshold | NOT SUPPORTED |

| | | | | |
|-----|--|---|---|-----------|
| H4 | XGBoost has the lowest MSE (excluding Decision Tree overfitting) | Ranking of MSE values across valid models | Lowest MSE model: XGBoost | SUPPORTED |
| H5 | If TBF > 9000, failure is likely | Proportion + t-test | t=4.814, p=0.000, 32.5% > 9000 | SUPPORTED |
| H6 | Downtime > 100 min is deviation | Z-score anomaly detection | 1 extreme events found | SUPPORTED |
| H7 | Failure sequence not informative | Linear regression R ² | R ² =0.0046 | SUPPORTED |
| H8 | Missing operational factors influence downtime | Indirect evidence | 3/7 models with R ² < 0.05 | SUPPORTED |
| H9 | Micro-damage accumulation leads to failure | Indirect evidence | Time-series RMSE ≈ 9653 | SUPPORTED |
| H10 | Degradation is non-linear/cyclic | Relative model performance | XGBoost R ² =0.9998 vs Linear R ² =0.0046 | SUPPORTED |

The evaluation of the tested hypotheses produced several insights into the underlying patterns of downtime behavior.

Inductive hypotheses H1 and H2 were supported, confirming that the sample means accurately reflect the underlying characteristics of the dataset. Nonetheless, their interpretive value remains limited due to the substantial variability observed in both time between failures and downtime duration.

Hypotheses H3 and H4, derived from analytical model outputs, produced mixed results. H3 was not supported, as only a portion of the machine-learning models exhibited very low explanatory power ($R^2 < 0.05$). H4 was supported once the overfitted Decision Tree model was excluded, confirming that XGBoost yielded the lowest mean squared error among the evaluated models. Hypothesis H7 was also supported, with the near-zero R^2 value indicating that the temporal sequence of failures contains no meaningful linear structure.

Deductive hypotheses H5 and H6 were validated through statistical testing. H5 showed that intervals exceeding 9,000 minutes represent a statistically significant deviation from the global mean and account for a notable share of events. H6 was confirmed through anomaly detection, which identified extreme downtime durations relevant for operational diagnostics.

Abductive hypotheses H8, H9, and H10 were supported through indirect evidence, including large residual errors in forecasting models, high variability in time series behavior, and inconsistencies across analytical methods. These findings suggest that additional operational, environmental, or degradation-related factors—absent from the recorded dataset—likely influence the occurrence of unplanned downtime. Collectively, the results demonstrate that the analytical framework provides a structured basis for interpreting failure patterns and generating meaningful hypotheses for further investigation.

5. Discussion

The results of the analysis demonstrate that the proposed analytical framework effectively supports the systematic generation and evaluation of hypotheses related to un-planned downtimes.

Beyond its conceptual contribution, the framework offers substantial practical value for researchers and practitioners. A complete Python implementation of all methodological steps—ranging from data preprocessing to hypothesis testing—is provided in the Supplementary File accompanying this article. Users only need to prepare a Data.xlsx file in the format specified in the supplementary materials. The code automatically executes all analytical stages and reproduces the full hypothesis-generation workflow exactly as demonstrated in this study. It is important to note that the present analysis filtered a single downtime category (tool wear); however, the same procedure can be applied to any other type of unplanned downtime simply by supplying a corresponding Data.xlsx dataset. In this way, the framework serves as a reusable and extensible tool that supports systematic exploration of failure patterns across diverse operational contexts..

A key limitation of the analytical framework lies in its dependence on recorded downtime data. The model requires precise time-stamped entries—specifically the exact hour and minute of the start and end of each downtime event, linked to the corresponding date and work shift—in order to calculate failure intervals and identify temporal patterns. The predictive capability of statistical and machine-learning models is further constrained by the number of collected events: small datasets make pattern detection difficult, increase output variability, and elevate the risk of overfitting. Consequently, the accuracy and generalizability of the generated hypotheses are inherently conditioned by the volume and precision of historical records of unplanned downtimes.

6. Conclusions

This study introduced and validated an analytical framework that enables structured hypothesis generation from historical downtime records. Rather than focusing on predictive accuracy alone, the framework emphasizes the interpretative value of quantitative analysis by guiding users toward formulating inductive, deductive, and abductive hypotheses that reflect different perspectives on failure mechanisms.

The empirical application demonstrated that even when predictive signals are weak or inconsistent, the framework can still extract meaningful operational patterns and highlight areas where additional contextual information is required. In doing so, it provides a systematic pathway for transforming raw analytical outputs into testable assumptions that can support deeper diagnostic or investigative work.

The approach is intended not as a standalone prediction tool, but as a methodological bridge between data analysis and reasoning-based exploration. This contributes to more transparent, explainable, and context-aware practices in the study of industrial failure behavior.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Data.xlsx.

Author Contributions: Conceptualization, N.P., M.M.; methodology, N.P., M.M.; software, N.P., M.M.; validation, N.P., M.M.; formal analysis, N.P., M.M.; investigation, N.P., M.M.; resources, N.P., M.M.; data curation, N.P., M.M.; writing—original draft preparation, N.P., M.M.; writing—review and editing, N.P., M.M.; visualization, N.P., M.M.; supervision, N.P., M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Science, Technological Development and Innovation of the RS according to the Agreement no. 451-03-137/2025-03/200105 from February 4 th 2025, and Eureka project No. 20495, Smart Toolbox for Mining's Resilience, Sustainability And Financial Performance Improvement-SMARTMEN.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request..

Acknowledgments: The authors express their gratitude to the industrial partner who provided access to the operational downtime records used in this research. The authors also thank the technical staff for their assistance in data collection and documentation preparation.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mobley, R.K. *An Introduction to Predictive Maintenance*, 2nd ed.; Butterworth-Heinemann: Oxford, UK, 2002; pp. 1–328.
2. Jardine, A.K.S.; Lin, D.; Banjevic, D.A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510.
3. Carvalho, T.P.; Soares, F.A.A.M.N.; Vita, R.; Francisco, R.P.; Basto, J.P.; Alcalá, S.G.S. A Systematic Literature Review of Machine Learning Methods Applied to Predictive Maintenance. *Comput. Ind. Eng.* **2019**, *137*, 106024.
4. Windmark, C.; Ståhl, J.-E.; Gabrielson, P.; Andersson, C. A Production Performance Analysis Regarding Downtime and Downtime Pattern. In Proceedings of the 22nd International Conference on Flexible Automation and Intelligent Manufacturing (FAIM 2012), Helsinki, Finland, 10–12 June 2012.
5. Heit, E. Properties of inductive reasoning, *Psychonomic Bulletin & Review* **2000**. vol. 7, no. 4. pp. 569–592. doi: 10.3758/BF03212996.
6. Coley, J.D.; Hayes, B.; Lawson, C.; Moloney, M. Knowledge, expectations, and inductive reasoning within conceptual hierarchies. *Cognition* **2004**, vol. 90, no. 3, pp. 217–253. doi: 10.1016/S0010-0277(03)00159-8.
7. E. Heit. Models of inductive reasoning. *The Cambridge handbook of computational psychology*. Cambridge University Press: New York, NY, US, 2008. pp. 322–338. doi: 10.1017/CBO9780511816772.015.
8. Hayes, B.K.; Heit, E. Inductive reasoning 2.0. *WIREs Cognitive Science* **2018**. vol. 9, no. 3, 13. <https://doi.org/10.1002/wcs.1459>.
9. Barrett, D.; Younas, A. Induction, deduction and abduction. *Evidence-Based Nursing* **2024**. vol. 27, no. 1. pp. 6–7. doi: 10.1136/ebnurs-2023-103873.
10. Heit, E; Rotello, C.M. Relations between inductive reasoning and deductive reasoning, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **2010**, vol. 36, no. 3, pp. 805–812. doi: 10.1037/a0018784.
11. Shanahan, M. Prediction is deduction but explanation is abduction. *Proceedings of the 11th international joint conference on Artificial intelligence - Volume 2* 1989. San Francisco, CA, USA. pp. 1055–1060.
12. Hobbs, J.R.; Stickel, M.E.; Appelt, D.E.; Martin, P. Interpretation as abduction. *Artificial Intelligence* **1993**. vol. 63, no. 1, pp. 69–142. doi: 10.1016/0004-3702(93)90015-4.
13. Kovács, G.; Spens, K.M. Abductive reasoning in logistics research. *International Journal of Physical Distribution & Logistics Management* **2005**. vol. 35, no. 2. pp. 132–144. doi: 10.1108/09600030510590318.
14. Jurado, S.; Nebot, À.; Mugica, F.; Mihaylov, M. Fuzzy inductive reasoning forecasting strategies able to cope with missing data: A smart grid application. *Applied Soft Computing* **2017**. vol. 51. pp. 225–238. doi: 10.1016/j.asoc.2016.11.040.
15. Lu, S.C.Y.; Zhang, G. A combined inductive learning and experimental design approach to manufacturing operation planning. *Journal of Manufacturing Systems* **1990**. vol. 9, no. 2. pp. 103–115. doi: 10.1016/0278-6125(90)90025-D.
16. Tóth, P.; Pogatsnik, M. Advancement of inductive reasoning of engineering students. *Hungarian Educational Research Journal* **2022**. vol. 13, no.1. pp 86-106. doi: 10.1556/063.2022.00120.
17. Franke, F.; Franke, S.; Riedel, R. AI-based Improvement of Decision-makers' Knowledge in Production Planning and Control. *IFAC-PapersOnLine* **2022**. vol. 55, no. 10. pp. 2240–2245. doi: 10.1016/j.ifacol.2022.10.041.
18. J. López Herrera. Time series prediction using inductive reasoning techniques. Doctoral thesis, Universitat Politècnica de Catalunya (UPC), 2000.

19. Filipič, B.; Junkar, M. Using inductive machine learning to support decision making in machining processes. *Computers in Industry* **2000**. Varna, Bulgaria. vol. 43, no. 1. pp. 31–41. doi: 10.1016/S0166-3615(00)00056-7.
20. Nikitenko, A. Robot Control Using Inductive, Deductive and Case Based Reasoning. *KDS 2005 conference proceedings* **2005**. Varna, Bulgaria. pp. 418–427.
21. Ou, J.; Wu, J.; Liu, C.; Zhang, F.; Zhang, D.; Gai, K. Inductive-Deductive Strategy Reuse for Multi-Turn Instructional Dialogues. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* **2024**. Miami, Florida, USA. pp. 17402–17431. doi: 10.18653/v1/2024.emnlp-main.964.
22. Grek, Å.; Hartwig, F.; Dougherty, M. An Inductive Approach to Quantitative Methodology – Application of Novel Penalising Models in a Case Study of Target Debt Level in Swedish Listed Companies. *Journal of Risk and Financial Management* **2024**. vol. 17, no. 5. p. 207. doi: 10.3390/jrfm17050207.
23. Vyas, N.; Woodside, A.G. An Inductive Model of Industrial Supplier Choice Processes. *Journal of Marketing* **1984**. vol. 48, no. 1. pp. 30–45. doi: 10.1177/002224298404800103.
24. Kaufmann, L.; Wagner, C.M.; Carter, C.R. Individual modes and patterns of rational and intuitive decision-making by purchasing managers. *Journal of Purchasing and Supply Management* **2017**. vol. 23, no. 2. pp. 82–93. doi: 10.1016/j.pursup.2016.09.001.
25. Medianovskiy, K.; Pietarinen, A.V. On Explainable AI and Abductive Inference. *Philosophies* **2022**. vol. 7, no. 2. p. 35. doi: 10.3390/philosophies7020035.
26. Teijeiro, T.; Félix, P. On the adoption of abductive reasoning for time series interpretation. *Artificial Intelligence* **2018**. vol. 262, pp. 163–188. doi: 10.1016/j.artint.2018.06.005.
27. Saylam, A.; Atlı, H. Predictive Analytics for Production Line Downtime: A Comprehensive Study Using Advanced Machine Learning Models. *The European Journal of Research and Development* **2023**. vol. 3, no. 4. pp. 88–94. doi: 10.56038/ejrd.v3i4.354.
28. Cummins, L.; Sommers, A.; Ramezani, S.B.; Mittal, S.; Jabour, J.; Seale, M.; Rahimi, S. Explainable Predictive Maintenance: A Survey of Current Methods, Challenges and Opportunities. *IEEE Access* **2024**. vol. 12. pp. 57574–57602. doi: 10.1109/ACCESS.2024.3391130.
29. Kadam, S.; Kadam, A.; Sayed, Z.F.; Priya, S.; Pardeshi, J.; Nemanwar, S. Factory Downtime Prediction Using Machine Learning Algorithms. *Journal of Emerging Technologies and Innovative Research (JETIR)* **2023**. vol. 10, no. 5. pp. i424-i430.
30. Abdaljalil, S.; Kurban, H.; Qaraq, K.; Serpedin, E. Theorem-of-Thought: A Multi-Agent Framework for Abductive, Deductive, and Inductive Reasoning in Language Models. *Proceedings of the 3rd Workshop on Towards Knowledgeable Foundation Models (KnowFM)*, Association for Computational Linguistics: Vienna, Austria. 2025. pp. 111–119. doi: 10.18653/v1/2025.knowllm-1.10.
31. McAbee, S.T.; Landis, R.S.; Burke, M.I. Inductive reasoning: The promise of big data. *Human Resource Management Review* **2017**. vol. 27, no. 2. pp. 277–290. doi: 10.1016/j.hrmr.2016.08.005.
32. Yu, H.; Chen, K.; Liu, Z.; Tu, H.; Li, A. Inductive relation prediction with information bottleneck. *Neurocomputing* **2024**. vol. 609. p. 128503. doi: 10.1016/j.neucom.2024.128503.
33. Upmeier zu Belzen, A.; Engelschalt, P.; Krüger, D. Modeling as Scientific Reasoning—The Role of Abductive Reasoning for Modeling Competence. *Education Sciences* **2021**. vol. 11, no. 9. p. 495. doi: 10.3390/educsci11090495.

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