USER GUIDE

This document guides the evaluation of the Knowledge-Based Digital Twin (KBDT) prototype, designed to detect and diagnose faults in the wind turbine pitch controller and drivetrain. Developed to support condition monitoring in wind farms, the system is undergoing incremental validation in terms of functionality, reliability, performance, and real-world applicability. It aids maintenance teams by converting simulated and real SCADA and vibration data into diagnostic insights and mitigation techniques, using interactive visualizations and rule-based interpretations grounded in simulation and technical literature.

Currently, the system focuses on faults related to encoders, tachometers, and actuators within the pitch controller. The data used in this first validation stage comes from a dynamic model representative of a 4.8 MW turbine and is tested against a dataset with similar nominal characteristics (OpenFAST-based). The dynamic model feeds the knowledge base of the expert system, which in turn performs inference over a real test dataset. This relationship between dynamic modeling, data-driven learning, and inference on real-world signals lays the foundation for a digital twin framework for fault detection and diagnosis.

To explore different usage scenarios, the current version enables specialists to simulate practical challenges commonly encountered in real wind farms. For example, it is possible to assess fault classification performance under limited data availability, reflecting the real-world rarity of fault conditions. It is also possible to isolate conditions where the actuator is actively engaged (e.g., above-rated wind speed), thus reducing the training set to a more meaningful subset of operational behavior. These use cases help the expert verify system robustness under constrained data scenarios.

Future releases will expand the diagnostic scope to include a 2.05 MW turbine pitch controller dataset and a complete fault diagnosis module for the drivetrain. This module will offer new test configurations, such as selecting input features from different domains (time, frequency, or hybrid), including vectorized image representations, and combining simulated and real laboratory vibration data. These upcoming capabilities will provide greater flexibility for domain experts to explore inference quality, validate system behavior, and simulate realistic operational constraints.

Throughout the validation process, domain experts are encouraged to evaluate the reliability of statistical analyses, the practical relevance of diagnostic and control outputs, and the alignment of system behavior with real turbine conditions. Feedback is captured directly in the application through dedicated input fields positioned at key interaction points.

SUMMARY

1. SYSTEM OVERVIEW

This section introduces the condition monitoring system developed under the digital twin paradigm. The system was built following a recursive knowledge engineering methodology, where each phase is developed incrementally. After each cycle, the outcomes are validated by a domain expert.

Two fundamental components are required to implement digital twins for condition monitoring: real operational data and a validated virtual model. These components were developed as follows:

- Model validation: A virtual wind turbine was developed using state-space equations and implemented within a multi-domain dynamic simulation platform. The model was calibrated and validated using real operational data, ensuring that it reproduces the turbine's behavior with high fidelity and reflects the physical dynamics of the real system.

- Knowledge-base: Once the dynamic model was validated, operational data under varying wind conditions were analyzed, including scenarios with and without sensor faults. This analysis supported the construction of the digital twin's knowledge base, forming the foundation for an intelligent inference engine capable of reasoning about system conditions and classifying potential faults.
- Inference mechanism: Leveraging extensive operational data encompassing various turbine conditions, multiple artificial intelligence models are employed to accurately classify faults, interpret signal patterns, and implement strategies to mitigate the impact of failures on system performance.

This initial validation stage is essential for understanding the turbine behavior in a controlled environment under a range of faulty and healthy operational conditions, before deploying the system with real-world datasets.

As a first test case, seven faults affecting encoders, tachometers, and actuators were introduced into the model-based wind turbine to generate scenarios for the knowledge-based. These fault scenarios were used to populate the knowledge base and guide the construction of a set of knowledge useful to diagnose the system's condition.

A hyperparameter-optimized artificial intelligence algorithm was employed to classify faults in a real dataset, based on simulated data from the knowledge base. Following the classification process, a PI controller is activated to mitigate the effects of sensor faults and restore the turbine normal performance.

The software also includes various KPIs that encompass correlation between twins and corrected and faulty signals, turbine performance, maintenance, and response time, both within the knowledge base and the real test datasets.

1.1. USER'S PROFILE

This system is designed for professionals involved in wind turbine performance and/or maintenance management. It aims to support users with varying degrees of expertise, particularly those engaged in:

- Maintenance management and performance monitoring of wind turbines;
- Condition-based and predictive maintenance;
- Asset management and KPI tracking;
- Wind turbine diagnostics and operations;
- Digital transformation initiatives (Industry 4.0) within the energy sector.

Data analysts and data scientists with strong backgrounds in statistical and artificial intelligence are also welcome to test the system and recommend improvements, since they also work in the wind energy sector.

Additionally, professionals working with dynamic modeling of wind turbines, data analysts and data scientists with strong backgrounds in statistics, artificial intelligence, or machine learning are encouraged to explore the system and contribute to its improvement, particularly in the context of wind energy applications. As the prototype relies heavily on sensors like encoders and tachometers, professionals with expertise in sensor technologies are also encouraged to contribute their insights.

Although the system has been designed to be user-friendly, it assumes a basic familiarity with the operational context and turbine performance metrics. Users are expected to understand the operational data available within the system, including sensor readings, actuators behaviors, vibration analysis, performance parameters, and failure indicators. A basic understanding of wind turbine failure modes and maintenance workflows is beneficial for making informed use of the results.

The system includes statistical visualizations, a confusion matrix, and performance scores. Users should be able to interpret classification outcomes and statistical summaries, particularly in the context of diagnostics. The system interface is fully interactive, allowing users to engage deeply with the data through powered visualization tools. Users can zoom in and out to inspect specific regions of interest, pan across plots to follow trends over time, and hover over individual data points to reveal precise values and contextual information. Additionally, the interface supports resetting views, selecting specific time windows or signal ranges, and downloading plots for offline analysis. Complementing this feature, the so-called "Expert Digital Twin Comments" serve as an assistive layer, offering contextual insights and interpretations related to the turbine's operational symptoms.

This documentation includes detailed instructions and background to help both newcomers and experienced maintenance managers effectively interpret system outputs and use them in support of decision-making.

2. SYSTEM TESTS

2.1. HOW TO ACCESS

The prototype of the conditional monitoring system can be accessed through the following link:

https://wt-kbdt.streamlit.app/

Streamlit is an open-source Python library that makes creating and sharing custom web applications for data analysis and machine learning easy. Although the library accepts aesthetic parameters of its functionalities through HTML, the code was developed 100% in Python, making the program expandable to any audience with basic knowledge of this programming language. The user does not need to install distributions or software that runs the Python programming language. Figure 1 shows a diagram of how the user-interface-data interaction works.

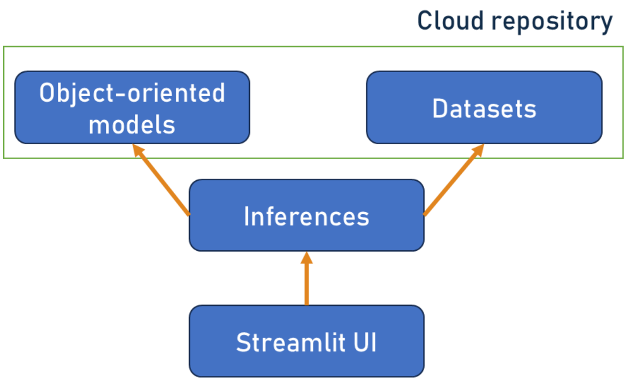


Figure 1. User interface (UI) with the prototype.

As illustrated in Figure 1, upon accessing the link, the user is directly connected to the prototype interface, which runs entirely in the cloud using Streamlit hosted environment. All inferences and data processing occur on Streamlit cloud infrastructure. The operational datasets and sensitivity analysis files are stored in a cloud repository and accessed dynamically by the app. Likewise, the application interface and code logic are also hosted in the same cloud environment.

Having established this foundation, once the user accesses the provided link, they are presented with Figure 2.

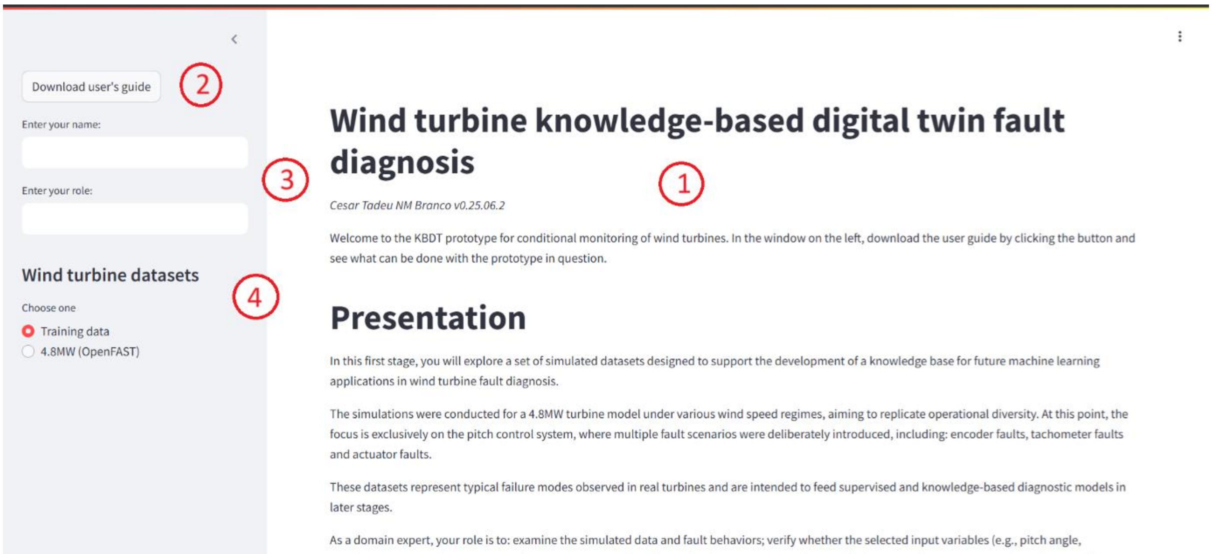


Figure 2. Initial page of the prototype.

Figure 2 presents the initial screen of the prototype. Item 1 refers to the main sliding analysis window. Items 2 to 4 correspond to elements of the sidebar. In item 2, the user can download the User Guide document. In item 3, there is a field where the user can personalize and name a ".csv" file that will store their feedback and diagnostic scores as they interact with the prototype. Throughout the prototype, there are blank fields followed by reflective questions designed to guide experts through the validation process.

In the sidebar, Item 4 provides access to a variety of datasets used to build the knowledge base and to test the inference capabilities of the artificial intelligence algorithms. These datasets are also intended to support the evaluation of the system diagnostic performance of the wind turbine.

The presentation logic of the two sections resides in an exploratory analysis of the data, automatic comments, performance indicators, and expert evaluation. This structure presents the data and offers strategic information to the users, in addition to collecting feedback. For now, only faults in the pitch controller are being presented. The drivetrain faults will be later implemented, as well as a second dataset for testing.

To streamline the presentation of test scenarios, the sidebar will be excluded from subsequent figures, allowing for greater clarity and emphasis on the core analytical components of the prototype.

2.2. DATA AND FAULT SCENARIOS

To validate the performance of the inference engine, specific faults were synthetically injected into the system sensitivity analysis datasets. These scenarios simulate typical pitch controller anomalies that might occur in real-world wind turbine operations, allowing the expert to evaluate the behavior of the turbine's variables under controlled conditions. Faults in sensors, encoders and tachometers were strategically injected into the signal by adding or multiplying the signal. Faults in actuators required changes in the controller parameters such as damping factor and natural frequency. The injected faults are summarized as follows:

- 1) Fixed value on pitch angle 1 equal to 5° in the time period from 2000 to 2100 s.
- 2) Gain factor on pitch angle 2 equal to 1.2 in the time period from 2300 to 2400 s.
- 3) Trend on the pitch angle 3, from 2° to 4° in the time period from 2600 to 2700 s.
- 4) Fixed value on rotor sensor speed equal to 1.4 rad/s in the time period from 1500 to 1600 s.
- 5) Gain factor on rotor sensor speed and generator sensor speed, respectively, equal to 1.1 and 0.9 in the time period from 1000 to 1100 s.
- 6) Change in the dynamics due to hydraulic pressure drop of the pitch actuator 2; the fault is assumed to be abrupt and it is present in the time period from 2900 to 3000 s.
- 7) Change in the dynamics due to increased air content in the oil on pitch actuator 3. The fault is slowly introduced during 30 s with a constant rate; afterward the fault is active during 40 s, and again decreases during 30 s. The fault begins at 3500 s and ends at 3600 s.

Both the "Training Data" section and the "4.8MW (OpenFAST)" turbine section contain all of these fault scenarios. However, the synthetic dataset in the first section was developed using a sensitivity analysis approach—that is, multiple datasets were generated considering different wind speed levels (5, 6, 8, 10, 12, 15, and 17 m/s), each with a variance of 1.5 m/s and a mean of 0 m/s.

The "4.8MW (OpenFAST)" turbine, in contrast, is based on a detailed simulation model developed by the National Renewable Energy Laboratory (NREL). This model is stored in the cloud and used for advanced diagnostic analyses and inference tasks, offering a more realistic representation of wind turbine dynamics.

2.3. TEST SCENARIOS

The tests provide support to the expert in decision-making, delivering statistical analyses, heuristic information, and metrics to assess the reliability of the diagnosis.

One of the innovative features of the digital twin framework is the automatic generation of expert-level insights, referred to as "Expert Digital Twin comments". These are textual interpretations generated based on the patterns identified in the plots. They translate technical observations into actionable insights, such as:

- Deviations from normal ranges
- Patterns suggestive of misconfiguration
- Sensor alignment or synchronization issues

2.3.1. Data analysis on the training data

The data analysis of the sensitivity analysis datasets serves as a strategic foundation for validating the system diagnostic capabilities. It provides a visual and quantitative baseline for assessing whether the digital twin realistically simulates turbine dynamics and identifies anomalies. By doing so, it supports the validation of automated inferences and expert comments generated by the system, ensuring they are consistent with field expectations and operational experience.

To enable this evaluation, the system allows users to explore multiple datasets representing distinct wind intensity profiles, each designed to expose the turbine to different operational regimes and fault conditions. These can be selected directly on the main interface, as shown in Figure 3.

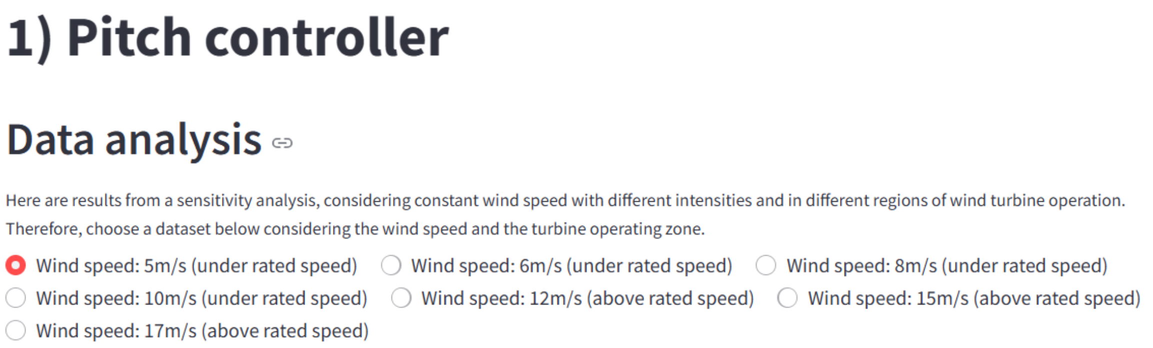


Figure 3. Different datasets from sensitivity analysis that can be instantiated in the conditional monitoring system.

To assess the behavior of the wind turbine under both normal and abnormal operating conditions, a set of statistical analyses was conducted using the main variables monitored by the system: wind speed, power generation, blade pitch angles, rotor speed, and generator speed.

The tools employed for this analysis included histograms, scatter plots, boxplots, and descriptive statistics. For wind speed, a histogram and scatter plot were used to evaluate the frequency and spread of recorded values, available in Figure 4.

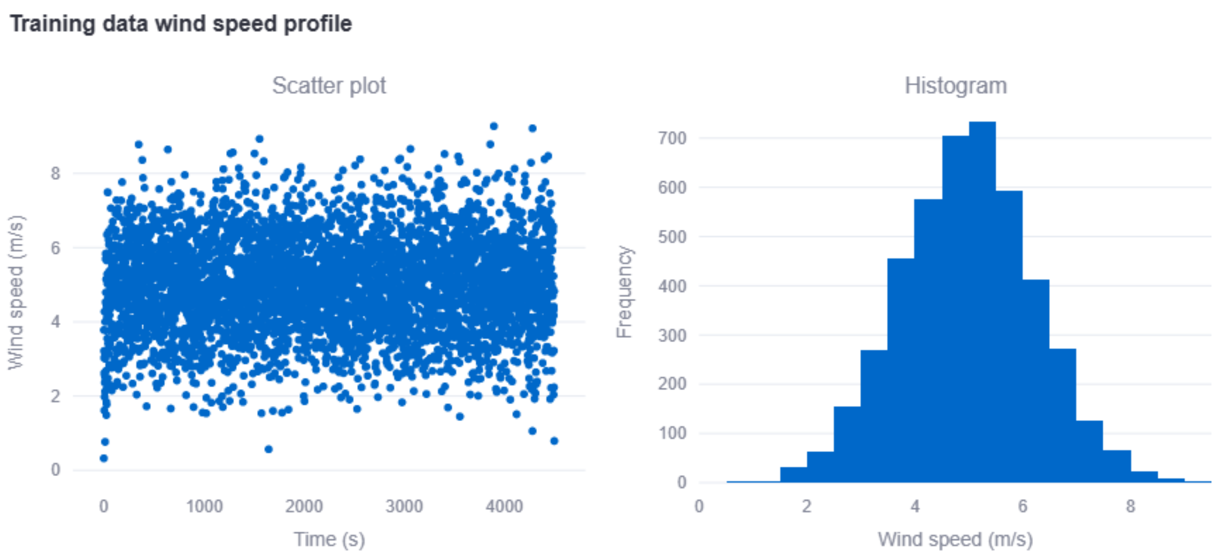


Figure 4. Wind speed scatter plot and histogram.

Figure 4 shows the user the dispersion of wind speed over time for the training set. The histogram shows the binomial behavior of wind speed and indicates the highest frequency of the class. For each dataset choice, the histogram will have binomial behavior with a mean around the chosen value of the dataset. Encoders signals for the three blades are displayed as scatter plots and boxplots (Figure 5).

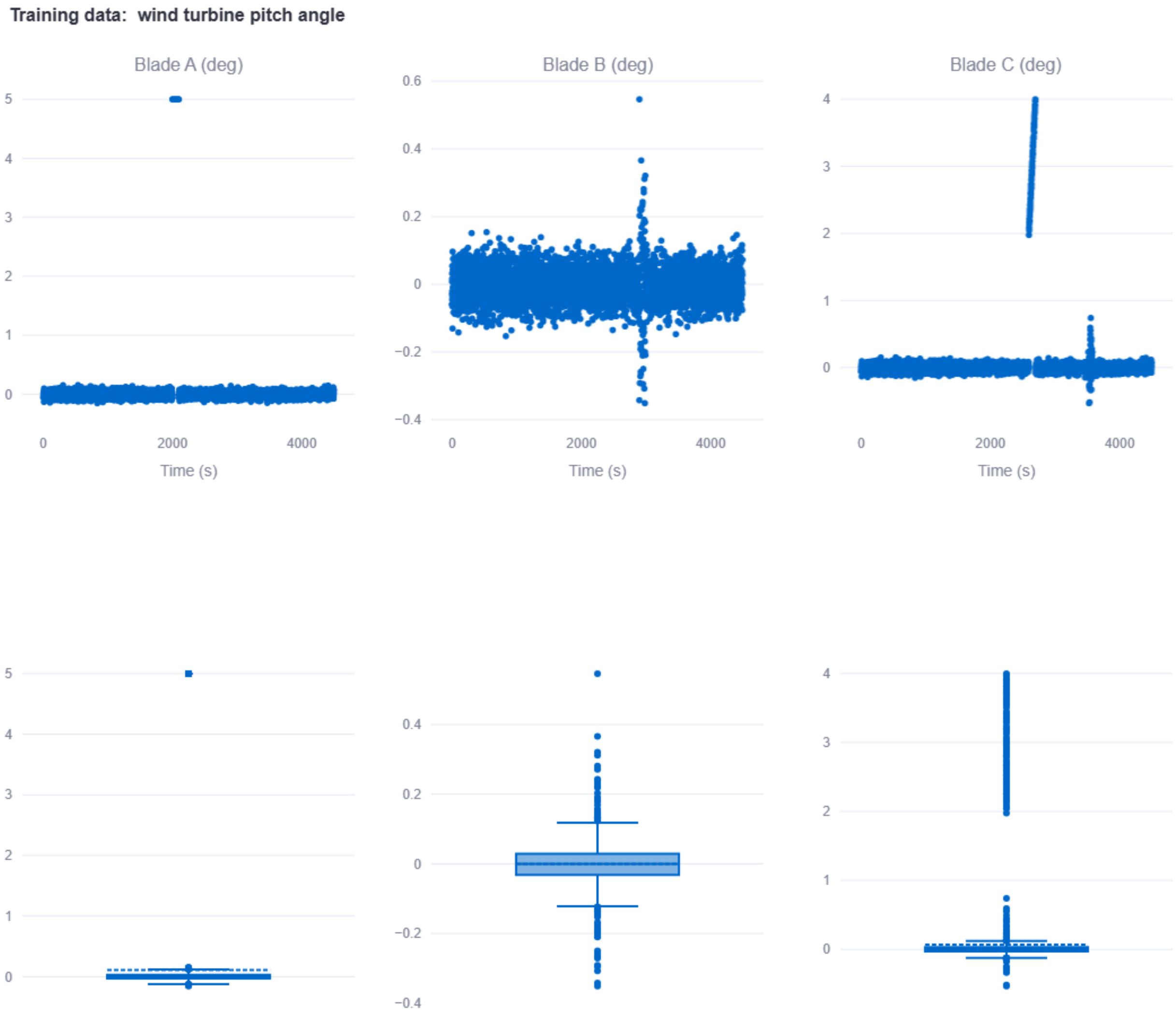


Figure 5. Scatter and boxplot of the blades.

Some points appear to be out of place in the scatter plot. These points are precisely faults injected at specific moments in the simulation. Details of the faults are specified in the prototype. This analysis already gives an idea of whether or not there is an anomaly in the blade sensor. The average angle of each blade is inside the boxes, followed by a comment from the "Expert Digital Twin". Furthermore, another scatter plot for rotor and generator tachometers are stated according to Figure 6.

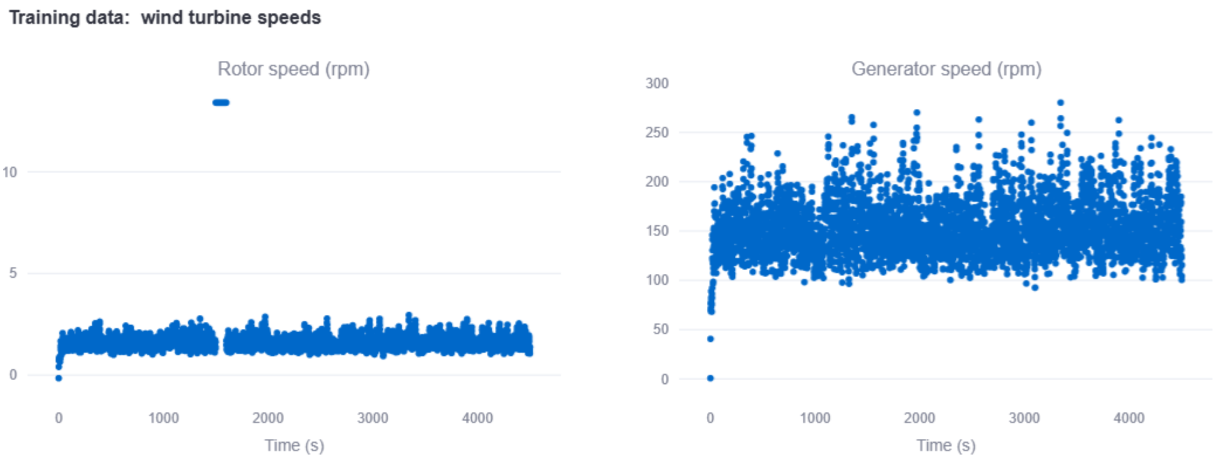


Figure 6. Rotor and generator speed scatter plots.

Again, there are extreme points that represent faults in the tachometers, in addition to the comments from the "Expert Digital Twin". The fault descriptions are presented in the prototype. To synthesize the relationships among all monitored variables, a correlation matrix was constructed according to Figure 7.

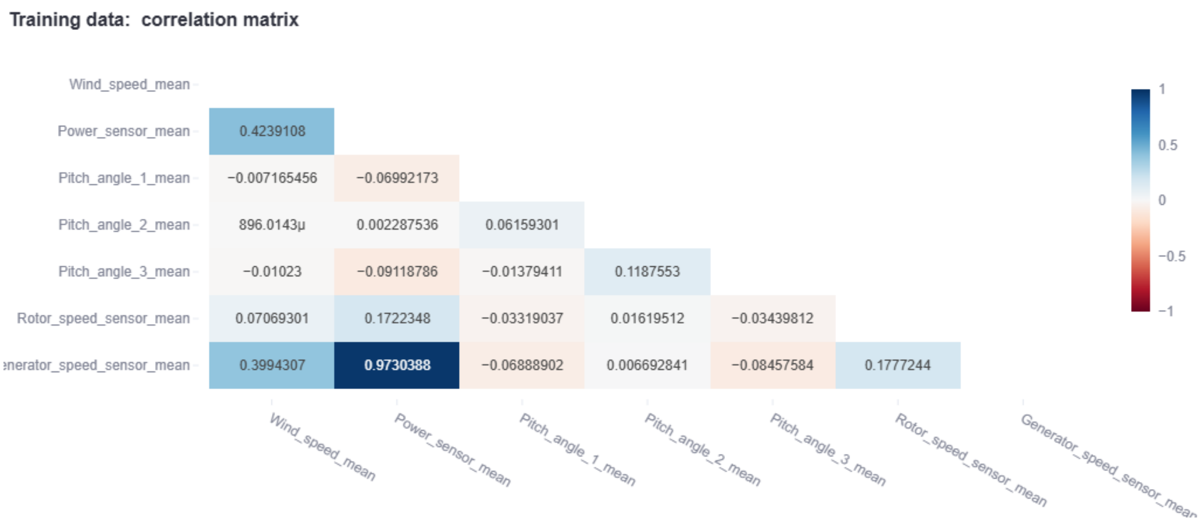


Figure 7. Correlation matrix among all variables.

This Figure 7 provides a high-level overview of how variables interact, helping to detect redundancies, strong dependencies, particularly useful for refining the knowledge base and improving fault inference logic.

2.3.1.1. Data analysis qualitative assessment

At the end of the statistical analysis, the expert is invited to leave his impressions based on a text to guide his writing.

How do you evaluate the usefulness of the statistical analyses provided in supporting diagnostic decision-making?
Expert comments:

In your opinion, do the profiles in the synthetic datasets reasonably reflect the behavior of turbine variables observed in real wind farms?
Expert comments:

Given the lack of real datasets with labeled faults, to what extent do you believe these synthetic datasets are suitable for training fault classification systems?
Expert comments:

2.3.2. System performance on the training set

In this section, the user is invited to evaluate how the digital twin system responds to different fault scenarios specifically affecting the pitch control mechanism. The goal is to analyze the operational impact of each fault and verify whether the system appropriately detects and reacts to abnormal conditions.

At the top of the section, the user can select different datasets. These include variations in wind speed, allowing experts to explore how the same fault may affect turbine performance differently depending on the operating context. The simulated faults are injected into the signals of sensors (encoders and tachometers) and actuators (pitch system), and each scenario is associated with a specific time window and failure pattern.

Instead of adjusting system parameters, the focus here is on interpreting results and understanding system behavior through a set of predefined Key Performance Indicators (KPIs) calculated for each fault:

- Total Energy (MWh): Indicates the total amount of energy produced during the simulation. It helps evaluate whether the fault caused any loss in energy generation.
- Capacity Factor (%): Represents the ratio between the actual energy output and the maximum possible output under ideal conditions. It is a measure of how efficiently the turbine was used.
- Full Load Hours (h): Refers to the number of hours the turbine would need to operate at full capacity to produce the same amount of energy generated during the simulation.
- MTBF (s) – Mean Time Between Failures: Reflects the average duration of operation between two fault detections, serving as an indicator of system reliability and fault recurrence.

These metrics allow the user to not only assess detection performance but also understand the broader implications of each failure mode on turbine efficiency and reliability.

This is not an exhaustive presentation of all the graphs generated by the system. Figure 8, therefore, serves as a representative sample of how the results of the sensitivity analysis are presented in the prototype.

Figure 8. Encoder with a fixed fault for an intensity of 17m/s.

The Figure 8 shows the variable "pitch angle" as a function of the wind speed for each turbine blade, accompanied by the power curve and the calculated key performance indicators (KPIs). Specifically, it illustrates the behavior of the rule-based diagnostic system when faced with a synthetic failure in the encoder of blade A, which remains stuck at 5° between 2000s and 2100s. The other blades operate normally to maintain constant turbine speed, as can be seen.

The signature of the failure is highlighted in red, which facilitates visual analysis. In this context, it is important for the specialist to assess whether the dynamic behavior of the turbine is consistent with that expected in the face of this type of failure and whether there is correspondence between the effects on the power curve and the KPIs presented.

In the next version, the system will incorporate a training dataset composed of vibration signals, dynamically modeled under different fault intensities. This version will form the second stage of the knowledge base, aimed at classifying faults in gearbox systems. In this phase, we will build a classifier using a diverse set of input features, including time-domain, frequency-domain, time-frequency features and vectorized image representations.

2.3.2.1. Training set qualitative assessment

At the end of the training set stage, the expert is invited to leave his impressions based on a text to guide his writing.

Do the threshold settings and binary alerts reflect realistic maintenance scenarios in wind turbines?
Expert comments:

Would these parameter configurations support accurate and timely fault identification in a real operational environment?
Expert comments:

Are the monitored variables (e.g., pitch angle, rotor speed) suitable for the type of failures being diagnosed? Please elaborate on your assessment.

Expert comments:

2.3.3. Multiclass fault classification

At sidebar, select the other dataset called “4.8MW (OpenFAST)”. In this phase, two activities are performed: one related to statistical anomaly detection, and the task of fault classification. In the first part of the section, there is a digital twin service responsible for detecting anomalies considering the power curve. Figure 9 shows this characteristic.

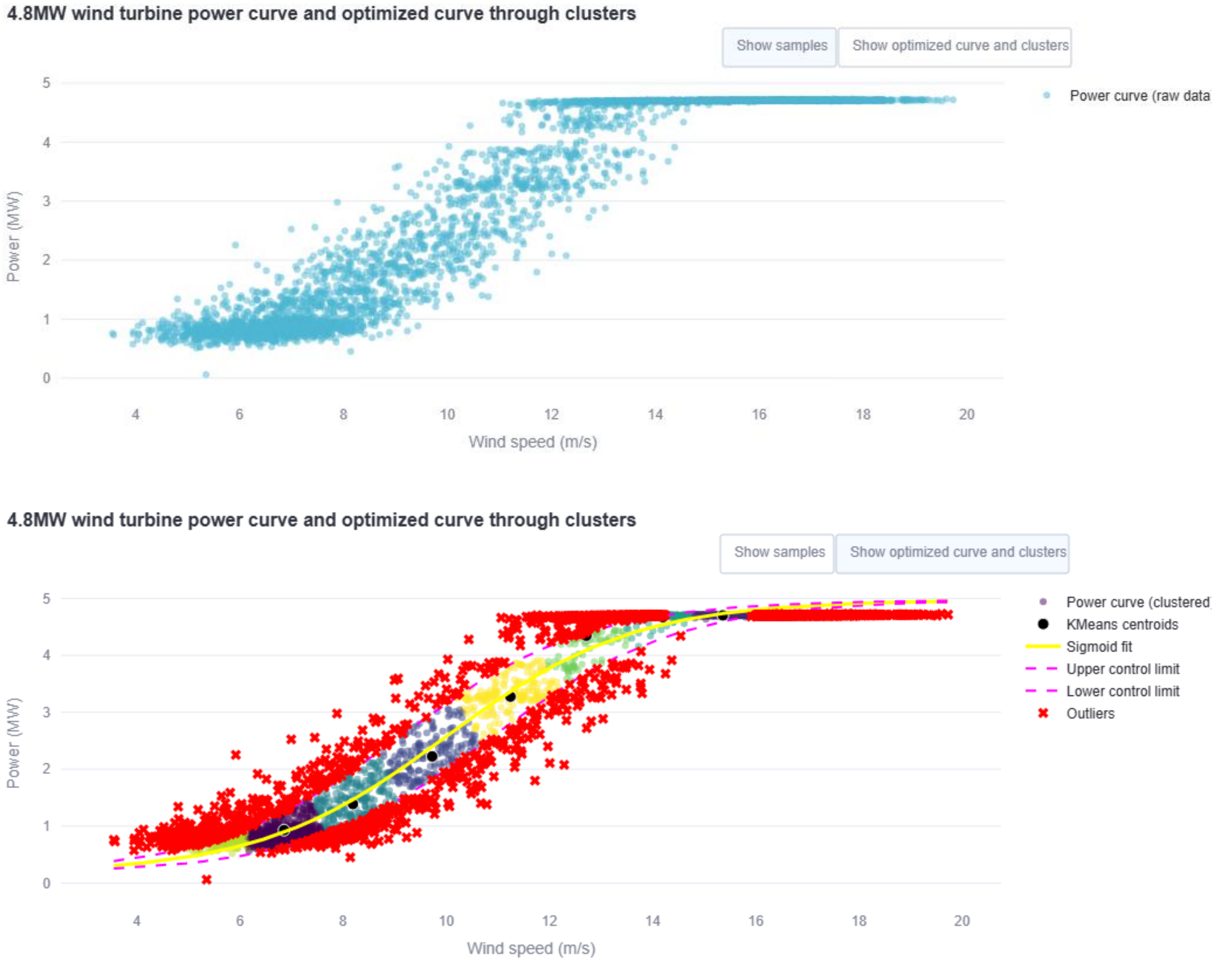


Figure 9. Wind turbine power curve generated by clusters.

In Figure 9, centroids are generated using K-means. With the generated centroids, a fitting curve based on the sigmoidal function is automatically adjusted to reproduce a replica of the real data. Control limits are implemented to classify what is normal and what is anomalous in the system. The data analysis over the real data and the correlation between are presented at the prototype. The user is invited to evaluate in the report itself the data analysis and insights into system anomalies carried out in the virtual twin inference.

In the Multiclass Classification section, the goal is to address a common and realistic challenge in fault diagnosis: the limited availability of labeled fault data in actual wind turbine operations. Though simulation allows us to generate as many fault scenarios as needed, real-world fault data are rare and difficult to obtain. To simulate this limitation, the system allows users to restrict the size of the training dataset using a slider control (as shown in Figure 10). This enables users to test the model performance both under full data availability (100%) and under more constrained scenarios.

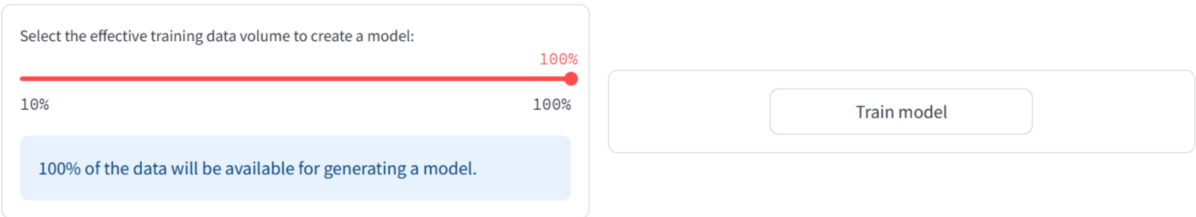


Figure 10. Slider to select the data volume inside the knowledge base and a button to train the model.

When the user selects a percentage lower than 100%, the reduction is applied proportionally across all classes. For example, selecting 40% will include 40% of healthy samples, 40% of fixed encoder fault samples, 40% of encoder gain fault samples, and so on. This ensures that the class distribution remains balanced relative to the reduced size while still emulating realistic data scarcity.

Once the dataset size is defined, the system proceeds to build a machine learning model using a fixed 80/20 split for training and validation, respectively. This process follows a hold-out validation strategy. A pre-selected ensemble learning model is used in this implementation, as multiple model architectures were tested offline and the most robust model was chosen for this prototype demonstration.

To generate the predictive model, simply click the “Train Model” button (see Figure 10). Once the training is complete, the system displays a “Classification Report” screen. This section presents the results of the fault classification stage, allowing the specialist to assess the performance of the machine learning model within the digital twin. The metrics shown include:

- Accuracy: Measures the overall correctness of the model. High accuracy indicates that the model is generally reliable in diagnosing the turbine condition.
- Precision: Shows how many of the predicted faults were actually correct. In the turbine context, high precision ensures that unnecessary maintenance or turbine shutdowns are avoided.
- Recall: Reflects the model ability to detect all real faults. This is especially critical for wind turbines, as a low recall (i.e., high false negatives) could mean actual faults go undetected, leading to increased wear, damage, or operational failure.
- F1-score: Balances precision and recall, providing an overall sense of the diagnostic system effectiveness.
- The report also includes macro and weighted averages, which help evaluate performance when fault samples are imbalanced — as is often the case in real wind turbine datasets.

The Confusion Matrix further visualizes the model predictions, showing how well it distinguishes between different fault types. Correct classifications appear along the diagonal. Off-diagonal values indicate confusion between fault classes or with healthy states, which is crucial to track when validating a diagnostic system’s robustness.

In addition to classifying faults, the system also demonstrates its ability to respond to them. Once a fault is identified, a fault-tolerant control mechanism is activated, which reconstructs the faulty signal using a PI-based approach. Each correction is evaluated through the following KPIs:

- Time Detection: The delay between when a fault occurs and when it is detected. Shorter detection times allow for faster intervention, minimizing risk to the turbine.
- Mean Absolute Error (MAE): Measures how closely the reconstructed signal follows the reference (healthy) signal. Lower MAE means better signal recovery and more stable turbine operation.

- Settling Time: The time taken for the reconstructed signal to stabilize after the correction begins. Shorter settling times indicate more efficient and effective recovery, reducing the operational impact of the fault.

This section gives the user insight not only into the detection but also into the resilience of the digital twin — evaluating how well it can restore normal operation after a fault, thus reinforcing the system's reliability in real-world turbine conditions.

2.3.3.1. Multiclass prediction qualitative assessment

Experts are now invited to critically assess the performance and operational utility of the multiclass classification approach. Please consider the ensemble-based confusion matrices in your evaluation, reflecting on their diagnostic value, reliability, and relevance to real-world turbine maintenance scenarios.

Does the system classification performance, as seen in the confusion matrices, KPIs, power curve, etc.., provide an accurate and operationally useful representation of fault and healthy conditions?
Expert comments:

In your experience, do the input variables and inference offer sufficient discriminatory power for real diagnostics?
Expert comments:

Would you recommend any improvements in the analysis above?
Expert comments:

3. QUANTITATIVE EVALUATION

At the end of each validation test, users are invited to provide qualitative feedback by answering open-ended questions tailored to each functionality of the system. These responses are essential to understanding the perceived value and practical relevance of the prototype from an expert's perspective.

In parallel, a quantitative assessment is also performed through a star rating system (from 1 to 5 stars) across the following ten core attributes:

Effectiveness of classification output: evaluates whether the system is correctly identifying and classifying faults, which is central to validating its diagnostic reliability.

Similarity of simulated failures to real-world conditions: checks how realistic and representative the simulated faults are when compared to actual failures observed in operational wind farms.

Clarity of diagnostic comments and visualizations: measures how clearly the system communicates its findings through automated comments and interactive visualizations.

Relevance of thresholds and parameters for diagnostics: assesses whether the thresholds and decision rules used in both rule-based and ML-driven diagnostics make sense in real-world applications.

Ease of navigation and interface clarity: Captures the user experience in terms of how intuitive and accessible the system interface is.

Usefulness for maintenance planning and prioritization: determines whether the outputs of the system effectively support planning and prioritizing maintenance actions.

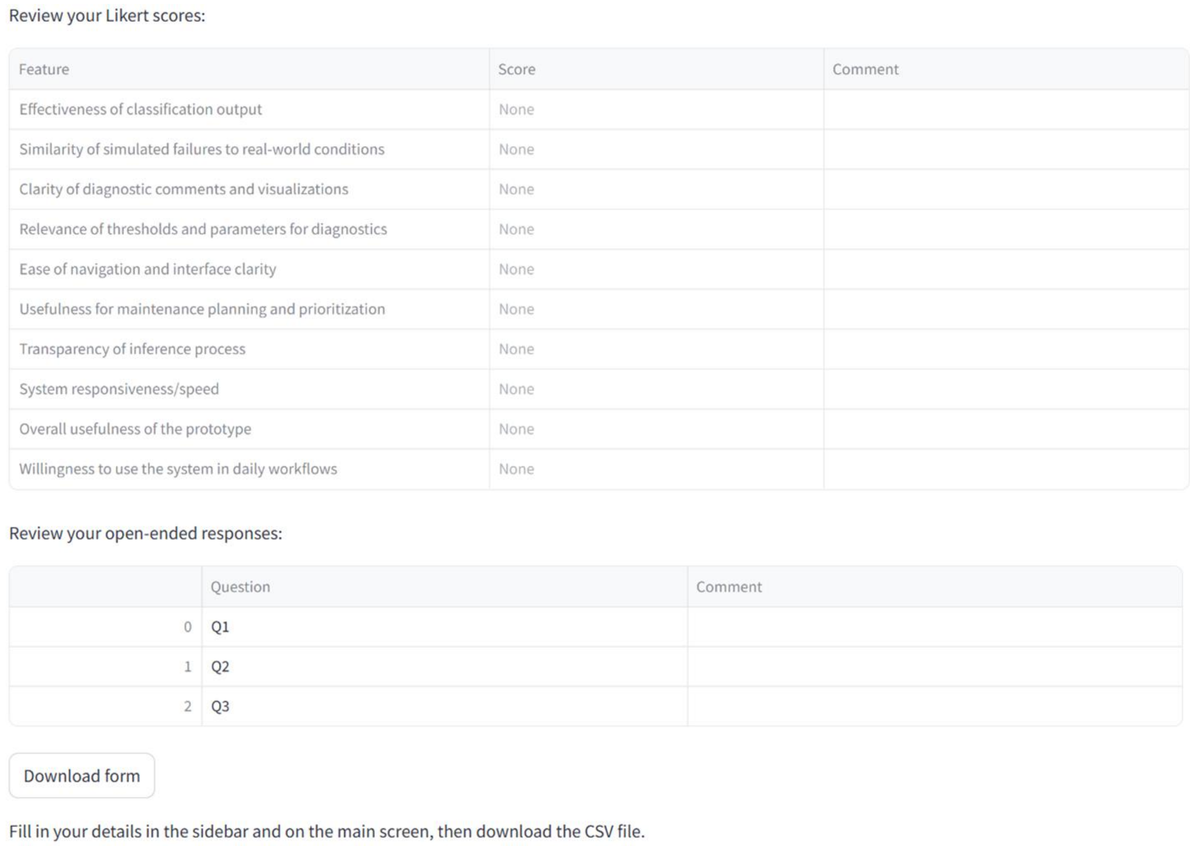
Transparency of inference process: evaluates how well the system explains the reasoning behind its conclusions, contributing to trust in its outputs.

System responsiveness/speed: assesses how quickly the system responds to user interactions and processes data, ensuring a smooth experience.

Overall usefulness of the prototype: assesses the overall value and contribution of the tool as a support system for condition monitoring and maintenance decision-making.

Willingness to use the system in daily workflows: measures the user's intention or likelihood to adopt the system in their routine work processes.

After evaluate the entire system, the user reaches a similar screen proposed on Figure 11.



Review your Likert scores:

Feature	Score	Comment
Effectiveness of classification output	None	
Similarity of simulated failures to real-world conditions	None	
Clarity of diagnostic comments and visualizations	None	
Relevance of thresholds and parameters for diagnostics	None	
Ease of navigation and interface clarity	None	
Usefulness for maintenance planning and prioritization	None	
Transparency of inference process	None	
System responsiveness/speed	None	
Overall usefulness of the prototype	None	
Willingness to use the system in daily workflows	None	

Review your open-ended responses:

	Question	Comment
0	Q1	
1	Q2	
2	Q3	

Download form

Fill in your details in the sidebar and on the main screen, then download the CSV file.

Figure 11. Review comments and scores screen and download form.

Figure 11 depicts the summary screen where all star ratings and qualitative comments are displayed for review. Users can revise and confirm their responses. Upon confirmation, the user is prompted to click “Download form”, which generates a downloadable Excel file containing all responses. If the user has provided their name and role, the file will be named in the format: “name_role_data.csv”. Otherwise, it will be saved as “unknown_user.csv”.

Participants are kindly requested to submit this file to the following email address: engmbcesar@gmail.com. To ensure your email is successfully delivered and not mistakenly flagged as spam, please also use an alternative method (e.g., send an email to Professor Jonny Silva or contact me at LinkedIn) to confirm that you have completed the validation process and submitted the file.