Intelligent Monitoring and Maintenance Prediction System for Industrial Equipment

Samuel Adetsi, Mu Ha, Cheng Zhang, Michael Hewlett

Table of contents

1 Executive Summary

Brilliant Automation, a leader in industrial automation solutions in Shanghai, China, has engaged our team to develop an intelligent monitoring and predictive maintenance system for industrial equipments. This project will leverage advanced data analytics and machine learning to transform sensor data into actionable insights, supporting early fault detection and optimized maintenance for key machinery.

2 Introduction

Brilliant Automation specializes in advanced monitoring and control systems for manufacturing and processing plants. To maximize equipment reliability and operational efficiency, the company installs high-frequency vibration and temperature sensors at critical points on key machinery, such as motors, gearboxes and bearing housings, across their clients' facilities. These sensors continuously collect data to monitor the health of essential assets including Tube Mills, Belt Conveyors, and High-Temperature Fans.

Example of sensor placement on industrial machinery

This project aims to enhance their predictive maintenance capabilities. By leveraging advanced data processing and machine learning, our system will enable early detection of equipment issues, reduce unplanned downtime, and optimize maintenance schedules. The focus of our analysis will be on three key pieces of industrial machinery: a Tube Mill, Belt Conveyor, and High-Temperature Fan. Our solution will integrate sensor exploratory data analysis, machine learning predictions, and an intuitive visualization platform to help maintenance teams make informed decisions about equipment upkeep and repair scheduling.

2.1 Context and Need

Manufacturing facilities face constant challenges in maintaining equipment reliability while minimizing maintenance costs. Current maintenance practices, which often rely on fixed schedules or reactive approaches, can lead to either unnecessary maintenance or unexpected breakdowns. Modern sensor technology and data analytics offer an opportunity to revolutionize this approach through data-driven decision making.

2.2 Core Challenges

Our project addresses several key challenges in industrial maintenance:

1. Converting complex sensor readings into meaningful maintenance indicators
2. Building transparent and reliable prediction models for equipment health evaluation
3. Creating an accessible interface for maintenance personnel

2.3 Key Goals

We aim to achieve the following:

1. Data Analysis and Understanding:
    - Map relationships across different sensor data
    - Identify patterns in equipment behavior
    - Analyze vibration signatures and their implications
2. Predictive Modeling:
    - Develop transparent prediction systems
    - Enable early fault detection
    - Provide clear reasoning for predictions
3. User Interface Development:
    - Enable live monitoring
    - Present clear status indicators
    - Facilitate historical analysis

2.4 Project Outputs

Our team will deliver the following key outputs:

- **Machine Learning Model for Device Ratings:**
  A robust machine learning model will be developed to analyze sensor and operational data, enabling the prediction and generation of health ratings for each monitored device. The model will be trained and validated using historical and real-time data, with performance metrics and retraining guidelines provided.
- **Interactive Dashboard for Visualization:**
  An intuitive dashboard will be created to display device health ratings and related analytics. The dashboard will support real-time monitoring, historical trend analysis, and customizable alerts, providing maintenance teams with actionable insights and a user-friendly interface for decision-making.
- **Comprehensive Final Report:**
  A detailed final report will be prepared for the client, documenting the project methodology, data analysis, model development process, results, and recommendations. The report will include visualizations, key findings, and guidance for future system enhancements or scaling.

3 Technical Approach

3.1 Data Overview

We have both input and output data from Apr 1 to Apr 15, 2025. The data consists of 3 devices which is summarized table below.

Table 1: Measurement System Overview

Equipment	Sensor_Points	Sensor_Data	Device_Ratings
Tube Mill	6 locations	5-second intervals	20-minute intervals
Belt Conveyor #8	4 locations	5-second intervals	20-minute intervals
High-Temperature Fan #1	5 locations	5-second intervals	20-minute intervals

3.1.1 Input Sensor Data

- Four key parameters are measured by the sensors at each location.
- These sensor readings are collected at 5-second intervals, providing high-resolution time series data for each piece of equipment.

Table 2: Input Data Summary

Sensor.Data	What.It.Does	Why.It.s.Important
Low Frequency Acceleration	Tracks slow vibrations	Detects alignment issues
High Frequency Acceleration	Tracks fast vibrations	Detect friction issues
Vibration Velocity Z (z-axis)	Tracks vibration strength vertically	Detect system damage
Temperature	Monitors component heat levels	Helps prevent overheating

3.1.2 Output Device Ratings

- The system generates 15 device health and status ratings, which serve as output parameters for each equipment unit. These include metrics such as RMS of Vibration Velocity, Crest Factor, Optimized Kurtosis, Rotor Balance Status, and others.
- Device ratings are produced every 20 minutes, summarizing the equipment's condition and performance based on the sensor data.
- These ratings are generated by a proprietary Matlab program running on the machines. The calculation process is a black box: even Brilliant Automation does not have access to the internal logic or algorithms used to derive these ratings.
- The ratings are out of 100. As for the definition of the ratings:

1. Above 80: Healthy
2. 60 to 79: Usable
3. 30 to 59: Warning
4. Below 30: Fault

Table 3: Device Output Rating Descriptions

Device.Rating	Description	Rating..0.100.
alignment_status	Alignment of conveyor components	0: Misaligned; 100: Perfectly aligned
bearing_lubrication	Lubrication level in bearings	0: Dry; 100: Fully lubricated
crest_factor	Ratio of peak amplitude to RMS value	0: Low peaks; 100: Severe peaks
electromagnetic_status	Condition of motor's electromagnetic field	0: Faulty field; 100: Stable field
fit_condition	Accuracy of component fit	0: Poor fit; 100: Perfect fit
kurtosis_opt	Kurtosis of optimized vibration signal	0: Low kurtosis; 100: High kurtosis
rms_10_25khz	Root mean square amplitude (10–25 kHz)	0: High amplitude; 100: Low amplitude
rms_1_10khz	Root mean square amplitude (1–10 kHz)	0: High amplitude; 100: Low amplitude
rotor_balance_status	Balance of the rotor	0: Imbalanced; 100: Perfect balance
rubbing_condition	Friction between components	0: Severe rubbing; 100: No rubbing
velocity_rms	Overall vibration severity	0: High vibrations; 100: Minimal
peak_value_opt	Optimized vibration peak value	0: Low peak; 100: Severe peak

3.2 Implementation Strategy

The data pipeline starts with sensor data stored in the client's internal database. An employee accesses this data using a remote desktop and copies it to their local computer. This step ensures that the original data is retrieved securely from the client's internal systems.

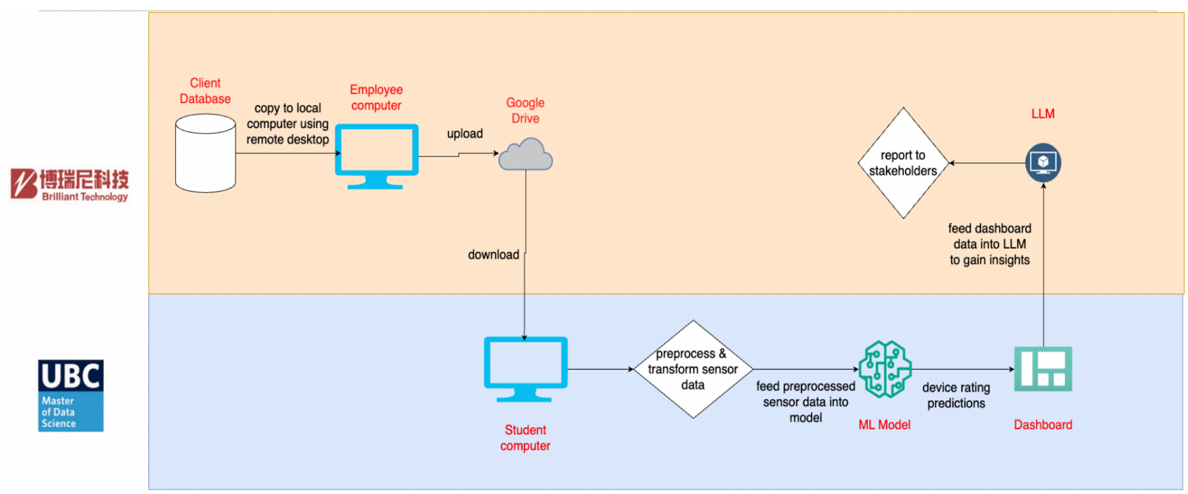
Next, the employee uploads the data to Google Drive. A student working on the project downloads the uploaded files from Google Drive to their own computer. This method allows for a smooth transfer of data between the company and the research team.

Once the student receives the data, they preprocess and transform it. This includes cleaning the data, selecting important variables, and reformatting it so that it can be used effectively by machine learning models.

After preprocessing, the sensor data is fed into a machine learning model. The model is trained to predict various device ratings, such as alignment status or vibration conditions. These predictions help assess the current state of the machines.

The model outputs are displayed in a dashboard. This dashboard provides a clear and interactive way to monitor machine performance over time. It helps both technical and non-technical users understand the system's health.

Finally, the data shown on the dashboard is sent into a large language model (LLM). The LLM analyzes the results and generates insights. These insights are used to create summary reports for stakeholders, making the results easier to understand and act on.



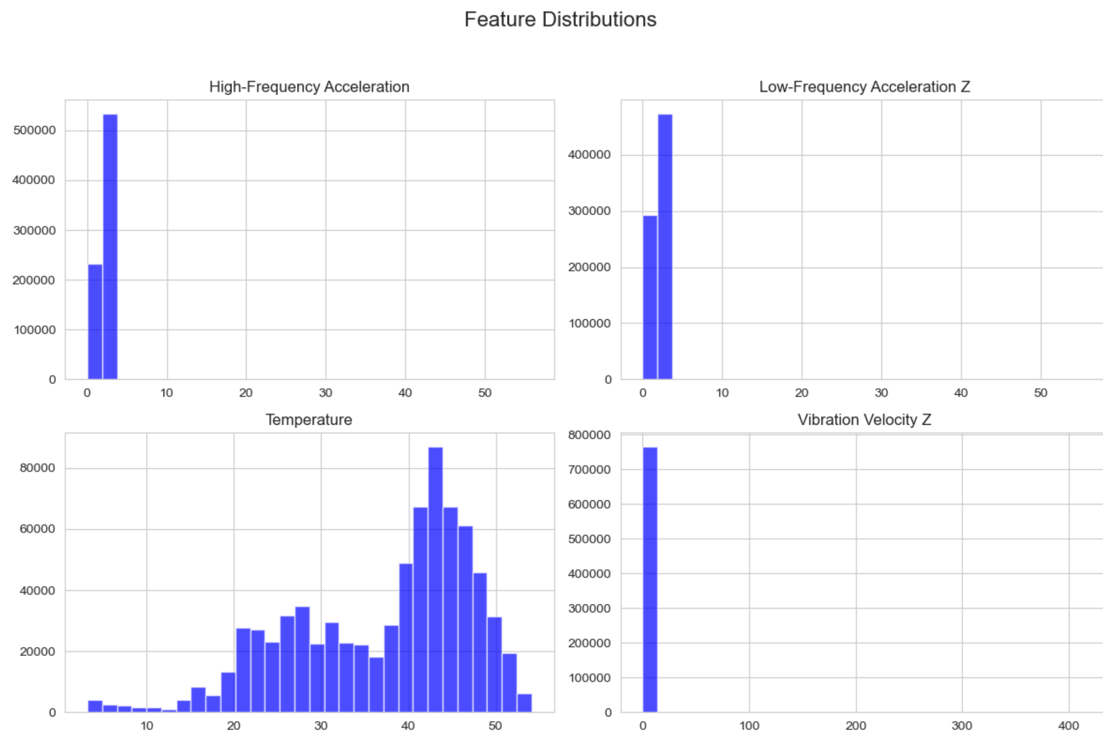
Overview of the end-to-end data pipeline

3.3 EDA and Data Processing

3.3.1 Input Features EDA

1. **Feature Distributions:**

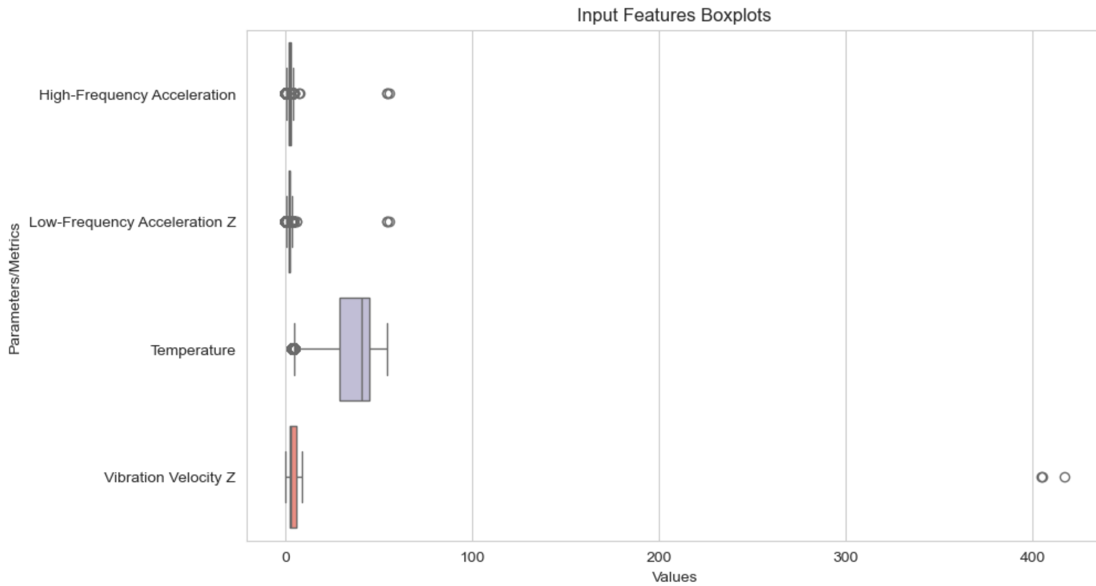
   - Visualizing the distributions of the input features helps assess their spread, skewness, and potential outliers.
   - This is essential for detecting feature scaling/normalization needs or discovering trends related to time dependencies.



Caption: Distribution of input features such as `High-Frequency Acceleration`, `Low-Frequency Acceleration Z`, `Temperature`, and `Vibration Velocity Z`.

2. **Boxplots for Sensor Parameters:**

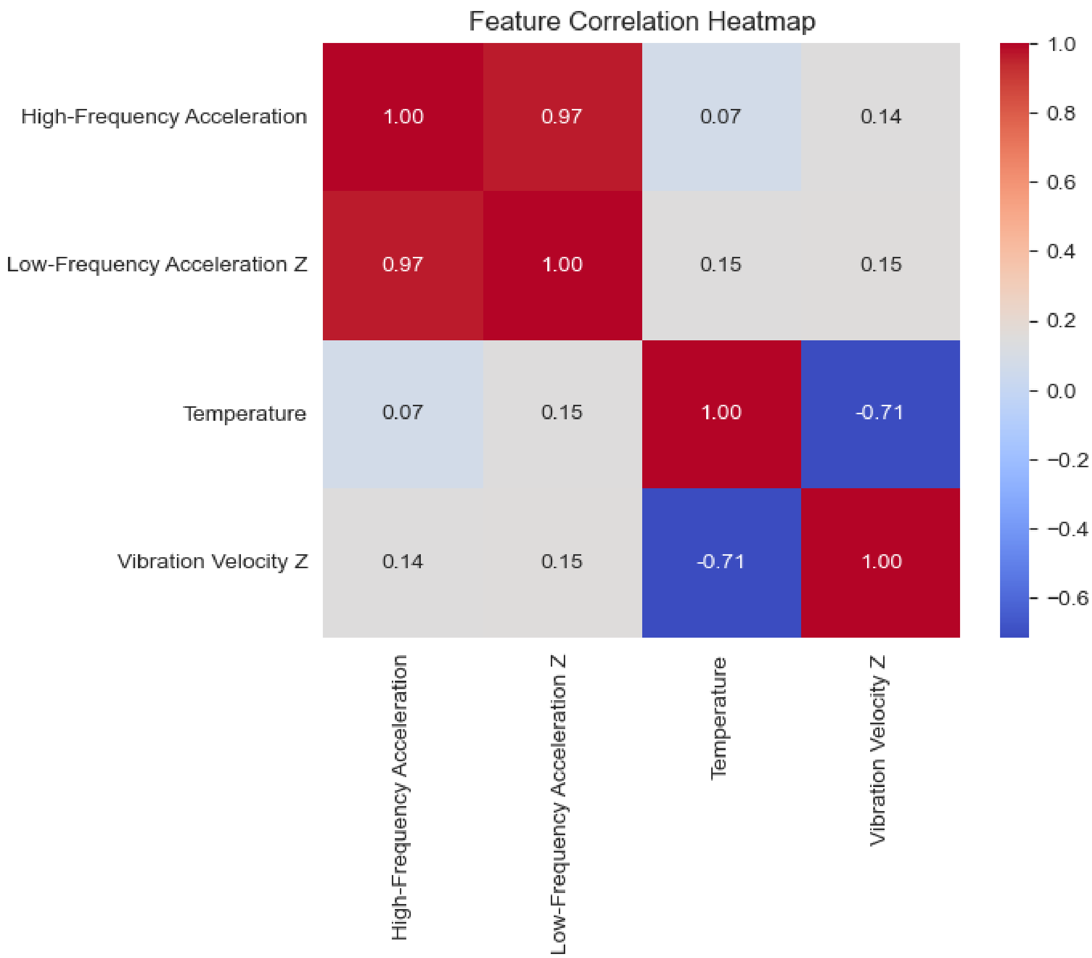
   - Boxplots show the range, quartiles, and potential outliers of input features.
   - This helps analyze anomalies or extreme measurements in parameters such as acceleration and vibration.



Caption: Boxplots for input features reveal the variability and presence of outliers.

3. **Feature Correlation Matrix:**

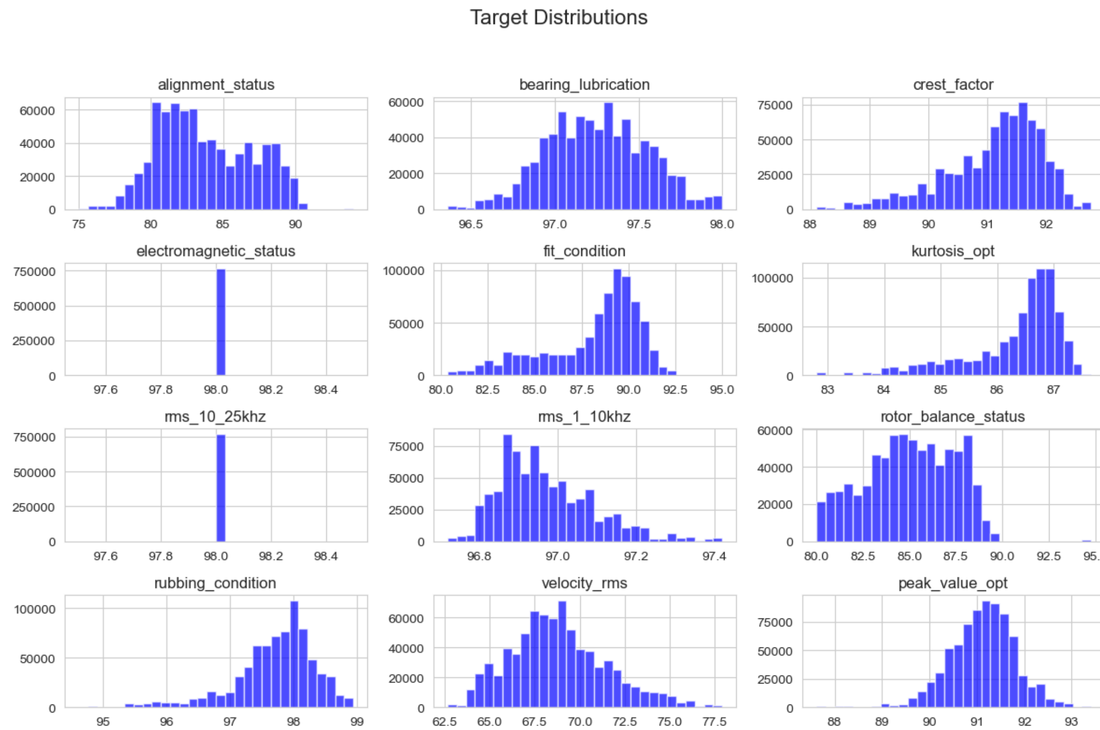
   - Assessing feature correlations helps identify strongly related parameters.
   - For example, higher vibration and acceleration might indicate worn-out components, making these features potential predictors of target ratings.



Caption: Correlation matrix of input features demonstrates the relationships between variables, important for feature engineering.

3.3.2 Target Features EDA

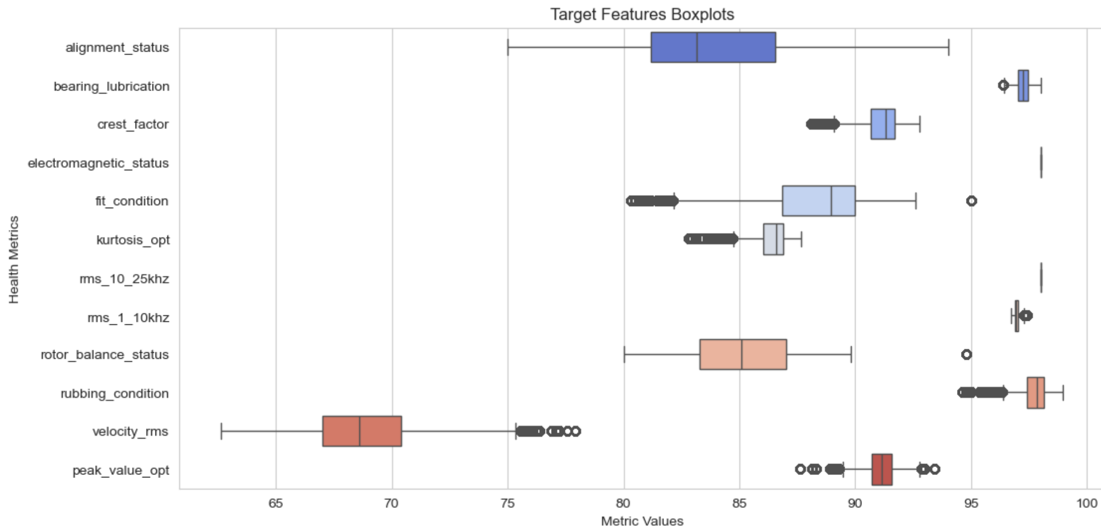
1. **Target Rating Distributions:**
    - Understanding the distributions of target ratings aids in identifying their variability and range.
    - Concentration in specific ranges might indicate clear thresholds for equipment health.



Caption: Distribution of target features such as `alignment_status`, `bearing_lubrication`, and other health metrics.

2. **Boxplots for Target Ratings:**

   - Boxplots for target features provide insights into their variability and detect potential inconsistencies (e.g., extreme high or low ratings could be indicative of data-quality or operational issues).



Caption: Boxplots for target features reveal rating variability and identify notable outliers in device health metrics.

3.3.3 Data Preprocessing:

We begin by combining the separate date and time columns into a single timestamp to simplify temporal analysis. The data is then pivoted so that each sensor measurement and device rating has its own column, creating a tidy format suitable for machine learning. Because the ratings are recorded every 20 minutes and sensor data every 5 seconds, each rating is duplicated across the corresponding 5-second intervals to align the data. For temperature values, which do not change significantly within short time spans, we use forward fill to handle missing values. This preserves data continuity without introducing significant bias. Sensor location and timestamp are included as features to capture spatial and temporal context. Lastly, the device column is dropped since we train a separate model for each device, making that column redundant.

3.4 Model Development

The model aims to predict device ratings for each equipment unit, using sensor data as input. Our modeling strategy progresses from simple to complex.

Table 4: Model Comparison Overview

Model	Complexity	Interpretability	Flexibility	Overfit_Risk	Compute_Cost
Baseline	Very Low	Perfectly clear	None	None	Minimal
Ridge	Low	High	Only linear fits	Low–Medium	Fast
PolyRidge (deg 2)	Medium	Medium	Simple non-linearities	Medium	Moderate
PolyRidge (deg 5)	High	Low	Highly flexible curves	High	Heavy

Random Forest	Medium–High	Low–Medium	Arbitrary non-linear	Medium	Moderate–High
Neural Network	High	Low	Very high	High	Heavy

3.5 Interactive Dashboard

The interactive dashboard serves as the central interface for maintenance teams and stakeholders to monitor machine health and understand sensor behavior in real time. It combines predictive model outputs and raw sensor readings in a clear, user-friendly layout. At the top of the dashboard, dropdown filters allow users to select a specific device and sensor, enabling targeted exploration. The radar charts visualize device health ratings across multiple metrics, helping teams quickly assess overall performance. Below and to the right, time-series and frequency plots show raw sensor data to help identify patterns, anomalies, or failure signals. This layered design allows users to connect machine learning predictions with actual sensor behavior.

The dashboard is built to meet industrial standards, as defined by client specifications. Its layout and visualization types are aligned with existing operational workflows, making it easy for technicians and analysts to interpret results. The responsiveness and modularity of the dashboard ensure it remains scalable for additional sensors or machines in the future.

4 Project Timeline

Table 5: Project Timeline and Outputs

Week	Stage	Outputs
1	Project launch, data processing	Wrangled dataset, toy dataset, MDS Proposal presentation
2	Data product MVPs	MVP dashboard, MVP models, MDS Proposal report
3	Full data test	Cloud computing pipeline, initial results
4	Model revision	Engineered features
5	Model revision	Engineered features
6	Output refinement	Final dashboard, final models, MDS draft data product
7	Output refinement	Final dashboard, final models, MDS presentation
8	Final checks	Final report, MDS final data product