Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Diagnostic Intelligence in 6G-Enabled Smart Factories: A Predictive Model for Machine Health and Operational Efficiency

Awinash Kumar

i.tnotes08@gmail.com

Jaipur National University https://orcid.org/0009-0002-6975-8154

Research Article

Keywords: Industrial IoT, Smart Factories, Predictive Maintenance, Data Engineering, Machine Health, Operational Efficiency, 6G Networks, Industry 4.0

Posted Date: August 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10621304/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

DIAGNOSTIC INTELLIGENCE IN 6G-ENABLED SMART FACTORIES: A Predictive Model for Machine Health and Operational Efficiency

Utilizing Python-based Data Engineering and Interactive Architectures for Proactive Industrial IoT Interventions at Thales Group

Author:
Awinash Kumar
Independent Researcher & Data Analyst

Project Access:

Live Streamlit App: https://6gsmartfactory.streamlit.app/

GitHub Code Repository: https://github.com/Awinashk567/Manufacturing-Process-Health-and-Operational-Efficiency-Analysis-in-6G-Enabled-Smart-Factories

Date: July 2026

Table of Contents

I. Abstract

II. Introduction

A. Background and Context

B. Problem Statement

C. Project Objectives

III. Analytical Methodology & Data Preprocessing

A. Dataset Description and Feature Selection

B. Data Validation and Handling Missing Values

C. Normalization of Physical Constraints

IV. Advanced Feature Engineering & Custom KPIs

A. Formulation of the Machine Health Index

B. Defect Density and Error Frequency Logic

C. Out-of-the-Box Concept: Shift-Based Efficiency Extraction

V. Exploratory Data Analysis (EDA) & Key Revelations

A. Threshold Warnings and Micro-Signals of Degradation

B. Cross-Metric Correlations: Vibration vs. Defect Rate

C. Underperformance and Production Stagnation

VI. Dashboard Architecture & UI/UX Integration

A. Interactive Control Panel and Filtering Logic

B. Visual Analytics Modules

VII. Technical Challenges & Strategic Problem Solving

A. Schema Mismatches and Hidden Whitespace

B. Pandas and Seaborn Future Warnings

VIII. Strategic Recommendations for Manufacturing Leadership

IX. Conclusion

X. Appendix A. Executive Summary for Government and Business Stakeholders

XI. References

I. Abstract

In the context of Industry 4.0, manufacturing processes are characterized by their continuous nature. However, monitoring hundreds of interconnected machines in search of inefficiencies and defects is an ongoing challenge for managers. This paper reflects on the development of a diagnostic intelligence system for Thales Group's 6G-enabled smart factories, which had to design keys performance indicators (KPIs) – Machine Health Index and Defect Density Score – to monitor and manage production. The results of the Streamlit dashboard, which processes industrial IoT data to make operational suggestions and predictions, were evaluated.

II. Introduction

A. Background and Context

Modern smart factories are complex systems comprising hundreds of interconnected machines, Industrial IoT (IIoT) sensors, and 6G networks with ultra-low latency. Industrial workers and their European defense counterparts, led by Thales Group, have one major concern in common: ensuring that production processes are of excellent quality. Before implementing powerful AI to predict equipment malfunctions, factory managers must first analyze the current state of technology and respond to critical questions: Are machines performing adequately? Where and why are they underperforming?

B. Problem Statement

Despite creating volumes of real-time sensor and production data, manufacturing teams still face the challenge of lacking central visibility over the health of their machines. By failing to gain essential insight into relationships between sensor behaviour, such as temperature and vibration, and production efficiency, manufacturers allow defects to slip through their quality control processes. The

inability to assess how individual machines perform compared to established standards results in poor-quality output and lower production rates.

C. Project Objectives

In this analytical project, my colleagues and I created a useful tool for manufacturing stakeholders. Our goal was to help them visualize and understand their data better. Here's what we were able to accomplish:

- Cleaning and preparing continuous IIoT sensor and production data
- Building mathematical models that estimate machine health and defect concentration
- Identifying machines that are at or near failure thresholds or have begun to deteriorate
- Developing a user-friendly Streamlit dashboard as an analytical control panel for factory managers.

III. Analytical Methodology & Data Preprocessing

A. Dataset Description and Feature Selection

The dataset contained real-time data from industrial machinery, including temperature, vibration, power consumption, production speed and network latency.

B. Data Validation and Handling Missing Values

A rigorous procedure was implemented to guarantee that all the data samples would be appropriate for the algorithm. It was found that more than 50,000 observations had inconsistent or missing entries for the Date and Timestamp fields. A program was written to combine these two fields into a datetime variable in a mathematical format that would be understandable by Python. Rows with faulty data could not be fixed and were subsequently removed.

C. Normalization of Physical Constraints

Industrial sensor readings have limitations to their physical values. Negative values in the power or vibration fields have no physical meaning and are typically caused by wrong sensor calibration. Absolute transformation of these fields was performed since negative values are impossible. In addition, missing sensor data was imputed by the corresponding machine's historical median to preserve the individual characteristics of a particular machine's behaviour rather than the general median across all machines.

IV. Advanced Feature Engineering & Custom KPIs

To elevate the analysis beyond basic averages, complex indices were engineered to accurately map machine behaviour.

A. Formulation of the Machine Health Index

A composite score was then developed to depict the overall health of the machine. This was done by normalizing Temperature_C , Vibration_Hz , and Power_Consumption_kW to their maximum levels respectively. A percentage value was then obtained by taking away the average of the normalized stresses from 100 as shown below:

$$Machine_Health_Index = 100 - \left(\frac{\frac{T}{T_{max}} + \frac{V}{V_{max}} + \frac{P}{P_{max}}}{3} \times 100 \right)$$

Where T, V, and P represent Temperature, Vibration, and Power Consumption, respectively. A higher score indicates optimal health.

B. Defect Density and Error Frequency Logic

To relate the number of defects to the amount of output, the Defect Density Score was developed. This score shows how many defective units were produced for every hour of production instead of using a percentage:

$$Defect_Density_Score = \left(\frac{Defect_Rate}{100}\right) \times Production_Speed$$

Additionally, the Error_Rate_% was mapped directly to an Error_Frequency_Index to track operational hiccups.

C. Out-of-the-Box Concept: Shift-Based Efficiency Extraction

The one concept that was truly out of the box was the extraction of "Shifts" of production data (Morning, Evening and Night) from the date time objects. Thus it becomes possible to analyze the changes in efficiency for different shifts rather than comparing individual machines.

V. Exploratory Data Analysis (EDA) & Key Revelations

A. Threshold Warnings and Micro-Signals of Degradation

By choosing 90% value as the threshold for temperature, an analysis was able to find 50 machines working under risky conditions. It allows the management to take appropriate measures to prevent serious accidents.

B. Cross-Metric Correlations: Vibration vs. Defect Rate

The scatter plots and the correlation matrix provided below help to see if there are any relationships between the mechanical stresses and the product. There is a positive correlation between the rise of Vibration_Hz and Error_Rate_%, therefore, the two variables are positively related. That implies that an increase in vibrations causes an increase in errors.

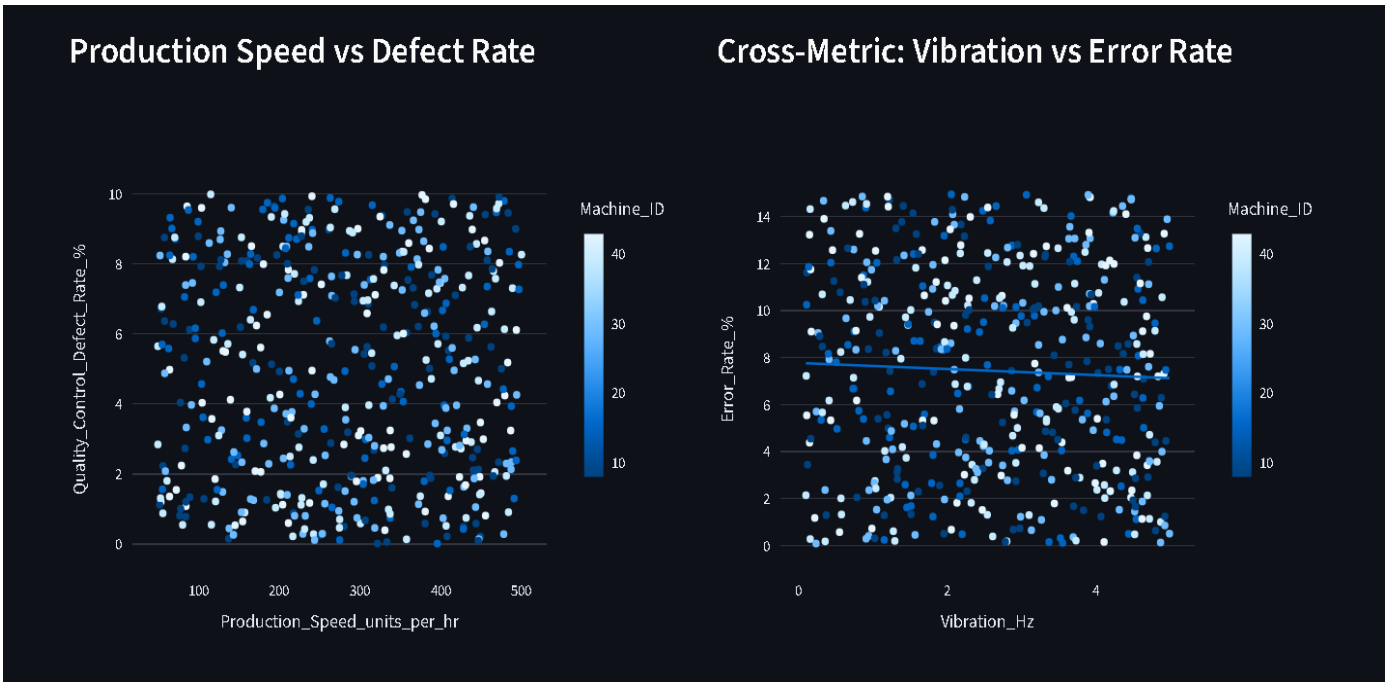


Figure 1: *Above presents scatter plots illustrating the correlation between several metrics, including the positive relationship between mechanical vibrations and the production error rate.*

C. Underperformance and Production Stagnation

The production performance diagnosis revealed that the bottom 10% of machines by Production_Speed_units_per_hr were underperforming. It indicates that these assets considerably impacted their production lines' Efficiency_Status, being the primary constraint.

VI. Dashboard Architecture & UI/UX Integration

A. Interactive Control Panel and Filtering Logic

The results of our analysis were implemented in the form of a web application, which was designed using the Streamlit framework. The application's sidebar on the lefthand side contains a Date Range picker, Operation Mode selector (as multi-toggle buttons for Idle, Active, Maintenance), and a Machine ID selector. The two latter filters enable

the floor manager to quickly narrow down the list of assets without performing any SQL queries.

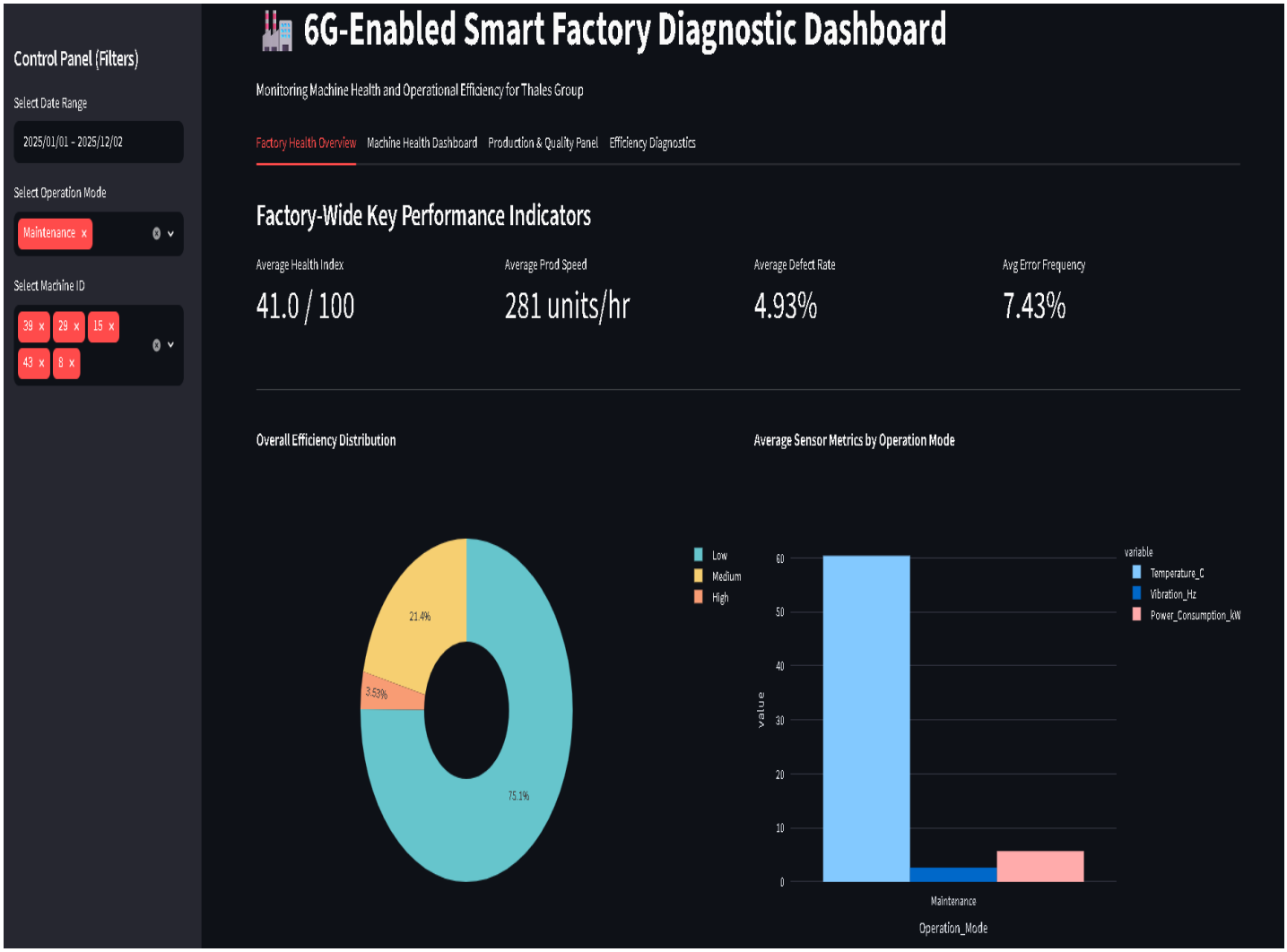


Figure 2: *The centralized Streamlit dashboard featuring the interactive Control Panel (filters) on the left and the real-time Factory Health Overview module.*

B. Visual Analytics Modules

The dashboard is structured into four distinct analytical tabs:

Factory Health Overview: Displays factory-wide averages for Health, Speed, Defect Rate, and an overarching Efficiency pie chart.

Machine Health Dashboard: Features a dynamic line chart for sensor trends over time, controlled by a user-driven metric toggle (Temperature, Vibration, Power), paired with a color-coded Health Scorecard.

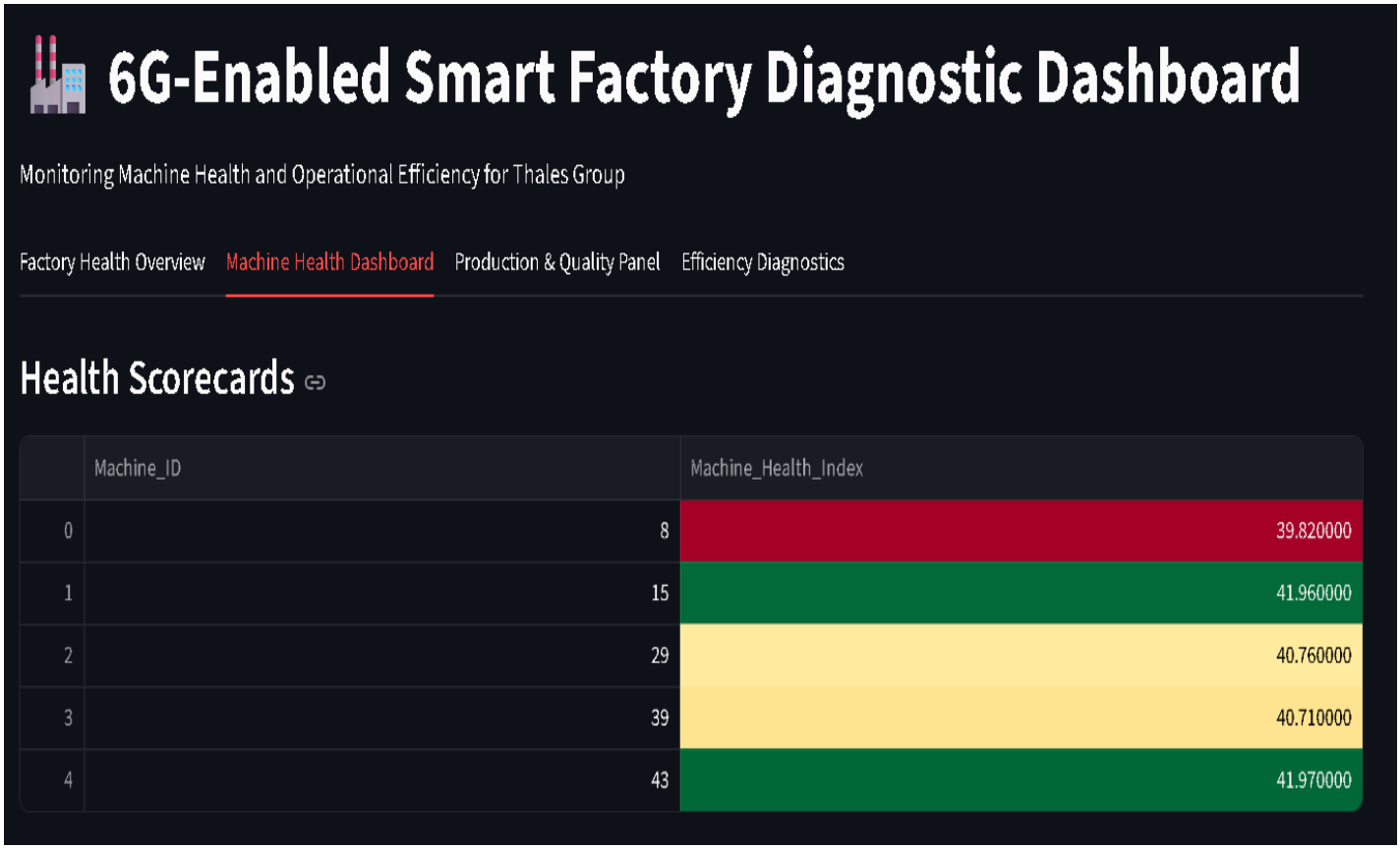


Figure 3: *The Machine Health Scorecard module utilizing conditional color-coding (Red/Yellow/Green) to enable managers to rapidly identify and prioritize degrading assets.*

Production & Quality Panel: Highlights scatter plots mapping Speed vs. Defect Rate, alongside dense time-series visualizations tracking Error Rate spikes.

Efficiency Diagnostics: Utilizes the engineered "Shift" feature to compare efficiency across different times of day, paired with boxplots examining power consumption trade-offs.

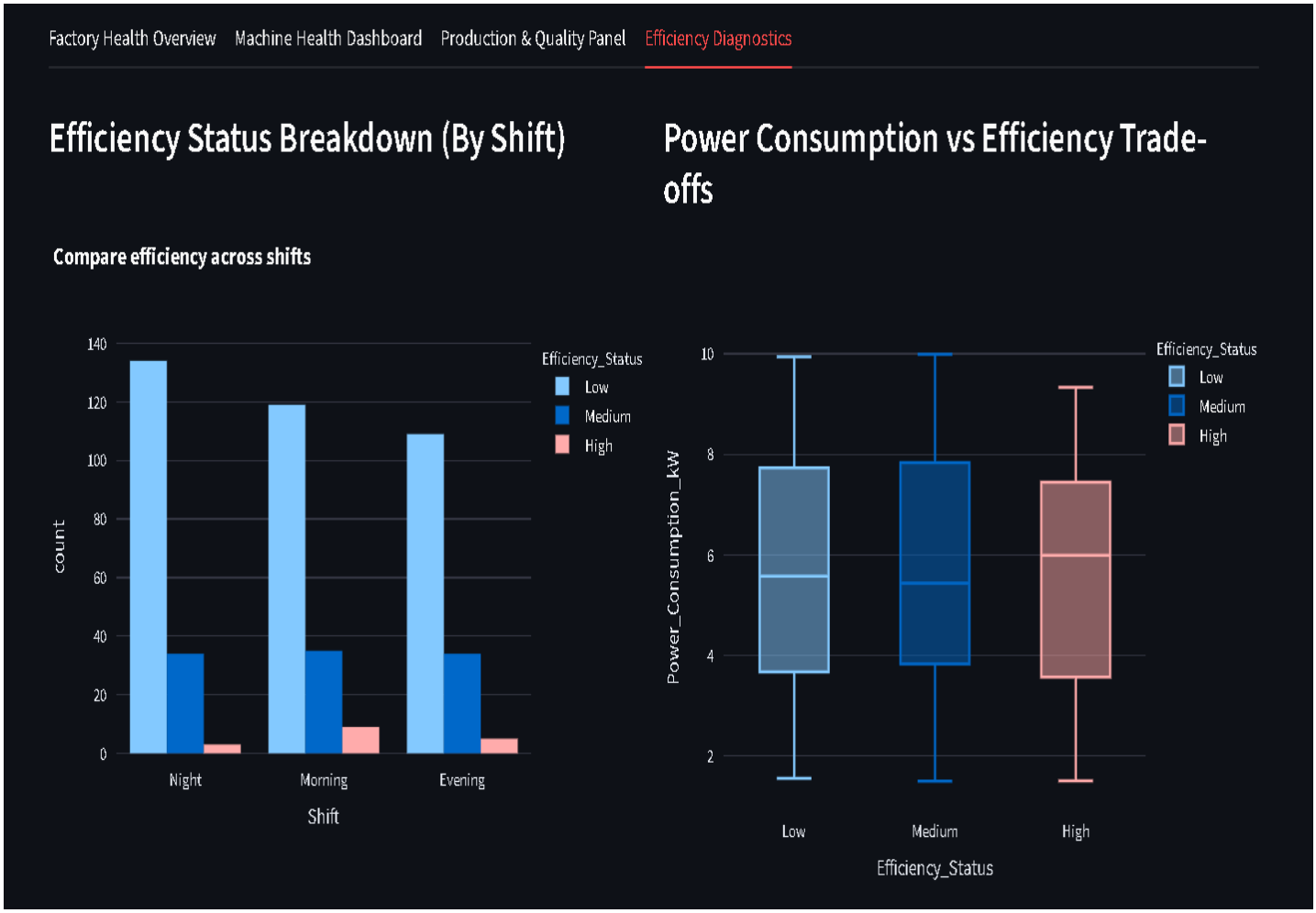


Figure 4: *Diagnostic view breaking down efficiency statuses across extracted operational shifts (Morning, Evening, Night) and mapping power consumption trade-offs.*

VII. Technical Challenges & Strategic Problem Solving

A. Schema Mismatches and Hidden Whitespace

Problem: Ingesting the .csv file resulted in a fatal Key Error: 'Time' when trying to create the datetime index at initialization.

Solution: Upon inspection, it was revealed that the column had a different name (Timestamp), and several columns include trailing/leading whitespaces from the data extraction process. A solution was implemented by performing .str.strip() on every header on a dynamic level, and the code was updated to reference the correct column nomenclature for proper ingestion.

B. Pandas and Seaborn Future Warnings

Problem: During the Exploratory Data Analysis (EDA), Seaborn module raised Future Warning messages about deprecation of plotting palettes using no hue variable.

Solution: To solve the problem of deprecation, the visualization code was updated to new standards (explicitly set hue and disable the legend). This will make the future implementation of the application warning-free.

VIII. Strategic Recommendations for Manufacturing Leadership

Based on the provided diagnostic results, the following process improvement suggestions are recommended for the Thales Group:

- Use the 90^{th} percentile threshold as a trigger for alerts in the 6G network to notify the technicians about temperature inconsistencies on a specific machine as soon as it approaches the critical mark.
- Base the maintenance schedule on the "Efficiency by Shift" visualizer to align the most destructive interventions with the least productive shifts (e.g., the Night shift).
- Prioritize the machines with abnormal vibrations for calibration rather than software-related maintenance since the former has shown its strong correlation with the number of defects.

IX. Conclusion

The proposed solution has shifted the initial concept of 6G-based smart factory management from maintenance-centric to diagnostic intelligence. It was achieved by transforming multiple IIoT data streams into an array of relevant Key Performance Indicators that can be easily evaluated using the Streamlit dashboard. Thus, the Thales Group will now be able to detect the most vulnerable machines in terms of

performance deterioration and set conditions for their autonomous AI-driven maintenance.

X. Appendix A. Executive Summary for Government and Business Stakeholders

Project Overview

For industrial giants such as the Thales group, ensuring uninterrupted production in 6 G-powered smart factories is essential to remaining competitive and meeting demand. Typically, production-related decisions are made in response to a malfunction or breakdown, which leads to inevitable downtime and loss of resources. The proposed project would utilize diagnostic intelligence to anticipate and troubleshoot malfunctions before they disrupt production and damage equipment.

Chief Challenges Identified

The analysis of raw information revealed some of the vital areas that were not previously considered in most cases. For instance, in several instances, the equipment was operating at around 90% of the limit without showing any significant alarm. In addition, it was discovered that the increase in the degree of vibrations resulted in a higher number of defects and errors during the production process. It would not be possible to determine the connection between the observed production rate, quality control, and sensor readings without considering the combination of these factors together.

Our Role (The Solution)

We engineered an end-to-end data pipeline and diagnostic system. From the raw data we filtered meaningful information, and through some involved calculations we produced several engineering KPI’s such as a composite Machine Health index as well as a defect density score among others. The results of which were showcased using

an intuitive interactive streamlit dashboard. This allowed the floor managers to slice and dice the data according to date, operating mode, and shift, and pinpoint the critical machineries affecting the production line.

Action Plan & Impact

By incorporating this dashboard into day-to-day operations, Thales Group can begin to take advantage of the system immediately. It allows managers to keep track of their health scorecards and provides an easy way of determining the point at which a machine needs calibration. As a result, defects will be minimized by having machines in good operating conditions, reducing breakdowns and, therefore, lowering overall costs associated with production. The implementation of this dashboard will facilitate the adoption of the coming generations of AI.

XI. REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," IEEE Network, vol. 34, no. 3, pp. 134-142, May/June 2020.

[2] L. D. Xu, E. L. Xu, and L. Li, "Industry 4.0: State of the art and future trends," International Journal of Production Research, vol. 56, no. 8, pp. 2941-2962, 2018.

[3] J. Lee, H. A. Kao, and S. Yang, "Service innovation and smart analytics for Industry 4.0 and big data environment," Procedia CIRP, vol. 16, pp. 3-8, 2014.

[4] W. McKinney, "Data Structures for Statistical Computing in Python," in Proceedings of the 9th Python in Science Conference, pp. 51-56, 2010.