Predictive Maintenance Using Engine Sensor Data
Interim Project Report

Siddhartha Mukherjee

January 31, 2026

Executive Summary

This interim report presents the development of a predictive maintenance solution using engine sensor data. The objective of the project is to anticipate engine failure events in advance by leveraging machine learning models trained on historical operational and sensor measurements.

During this phase, a reproducible data pipeline was established using the Hugging Face platform for dataset and model versioning. Exploratory data analysis revealed meaningful differences in sensor behavior between normal and failure conditions, as well as class imbalance in the target variable. These findings informed the selection of appropriate modeling strategies and evaluation metrics.

The dataset was cleaned, prepared, and split into training and testing subsets, which were registered back to the Hugging Face Dataset Hub to ensure traceability. Multiple machine learning models were trained and evaluated, with a Random Forest classifier selected as the best-performing model based on recall and F1-score.

The outcomes of this interim phase demonstrate the feasibility of predictive maintenance using the available sensor data and provide a strong foundation for further model refinement and deployment in the final phase of the project.

1   Introduction

Unplanned engine failures lead to operational downtime, increased maintenance costs, and reduced asset availability. Traditional reactive or schedule-based maintenance strategies are either costly or ineffective at preventing unexpected breakdowns.

Predictive maintenance leverages historical sensor data and machine learning models to anticipate failures before they occur. By identifying early warning patterns in engine operating conditions, maintenance activities can be scheduled proactively, reducing both downtime and cost.

This interim report focuses on the machine learning lifecycle up to experimentation tracking, including data registration, exploratory data analysis, data preparation, and model development.

2   Data Registration

A structured project directory was created to ensure a clear separation between raw data, processed datasets, and experimental artifacts. A dedicated `data/` subdirectory was used to store all datasets associated with this project, enabling organized data management and traceability.

To ensure reproducibility, auditability, and version control, the raw dataset was registered on the Hugging Face Dataset Hub. The dataset was uploaded in its original form prior to any preprocessing or transformation, ensuring that all downstream analysis can be traced back to a consistent and immutable data source.

All subsequent data loading operations in the analysis and modeling pipeline retrieve the dataset directly from the Hugging Face repository rather than from local file paths. This approach guarantees consistency across experiments and enables independent verification of results.

Dataset Details

- **Dataset Name:** predictive-maintenance-engine-data

- **Storage Platform:** Hugging Face Dataset Hub

- **Dataset URL:** `https://huggingface.co/datasets/mukherjee78/predictive-maintenance-engine-data`

- **Data Type:** Multivariate engine sensor data

- **Intended Use:** Predictive maintenance and engine failure prediction

3 Exploratory Data Analysis

This section presents an exploratory analysis of the engine sensor dataset to understand its structure, characteristics, and underlying patterns. The objective of this analysis is to gain insights into data quality, feature behavior, and relationships between variables, which will inform subsequent data preparation and modeling decisions.

3.1 Data Collection and Background

The dataset consists of multivariate sensor readings collected from industrial engines operating under varying conditions. Each engine is monitored over multiple operational cycles, with sensor measurements capturing different aspects of engine health and performance. The primary objective of the dataset is to enable predictive maintenance by identifying patterns that precede engine failure events.

Such sensor-driven monitoring systems are commonly used in industrial settings to reduce unplanned downtime and optimize maintenance schedules. By analyzing historical sensor behavior, early warning indicators of potential failures can be identified, allowing proactive intervention.

3.2 Data Overview

An initial structural assessment of the dataset was conducted to understand its scale, composition, and data quality. The dataset consists of multiple numerical variables representing engine operational parameters and sensor measurements recorded over time.

The dataset contains a single target variable indicating engine failure status, which is modeled as a binary classification problem. All remaining variables represent continuous sensor readings related to temperature, pressure, rotational speed, and fuel characteristics.

An inspection of the dataset revealed that the majority of features are numerical in nature, making the dataset well-suited for machine learning algorithms that operate on continuous inputs. A check for missing values confirmed that the dataset does not contain significant missing data, reducing the need for extensive imputation strategies.

Descriptive statistics were computed for all numerical features to examine their central tendency and variability. Several sensor variables exhibit wide value ranges, indicating differences in operational regimes and engine conditions. These observations motivate the use of robust models capable of handling heterogeneous feature scales.
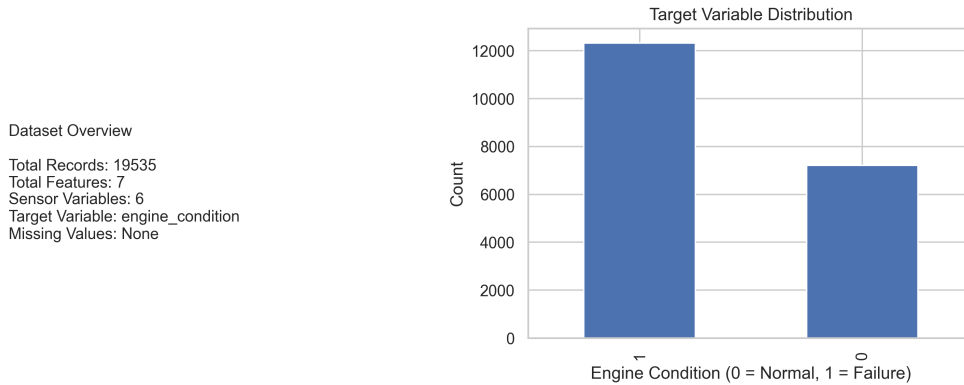

Dataset Overview

Total Records: 19535
Total Features: 7
Sensor Variables: 6
Target Variable: engine_condition
Missing Values: None

Figure 1: Overview of dataset structure and descriptive statistics

3.3 Univariate Analysis

Univariate analysis was performed to examine the distribution and variability of individual features. This analysis helps identify skewed distributions, potential outliers, and irregular sensor behavior.

The distribution of the target variable was analyzed to assess class balance. Additionally, representative sensor variables were visualized to understand their value ranges and overall behavior across engine cycles.
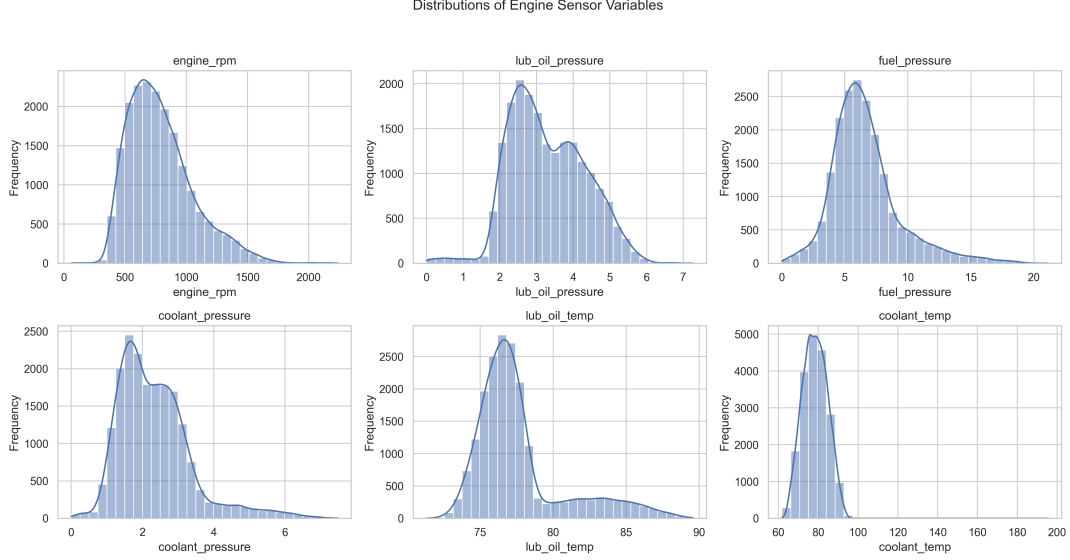

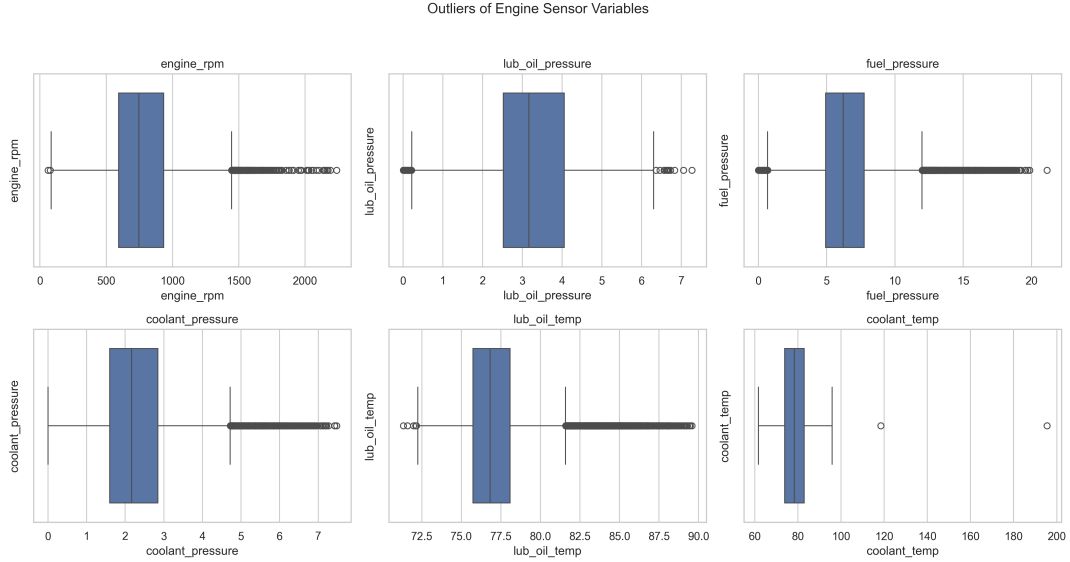Figure 2: Univariate distributions of key engine sensor variables

Figure 3: Outlier patterns observed across engine sensor variables

3.4 Bivariate Analysis

Bivariate analysis was performed to examine the relationship between individual sensor variables and the engine failure target. Sensor readings were compared across failure and non-failure instances to identify features that exhibit distinguishable behavior prior to failure events.

Boxplots were used to visualize the distribution of selected sensor variables across the two target classes. Several sensors show noticeable shifts in median values and increased variability in failure cases, suggesting that these variables may contain predictive signals related to engine degradation.

In addition to visual analysis, summary statistics grouped by failure status were examined to quantify differences in sensor behavior between the two classes. These comparisons highlight sensor variables that demonstrate systematic changes as engines approach failure.

The results of this analysis provide early intuition regarding feature relevance and support the selection of models that can capture non-linear relationships between sensor readings and failure outcomes.
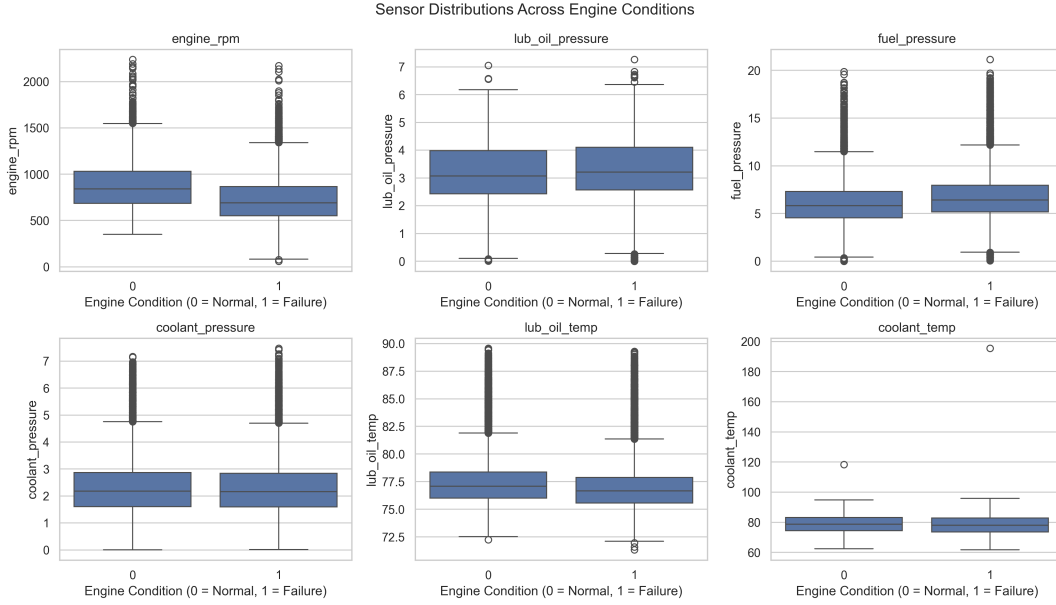

Figure 4: Sensor value comparison across normal and failure engine conditions

3.5 Multivariate Analysis

Multivariate analysis was conducted to explore relationships among multiple sensor variables simultaneously and to identify potential interdependencies. A correlation analysis was performed across sensor features to assess the degree of linear association between variables.

The correlation matrix reveals the presence of several strongly correlated sensor pairs, indicating redundancy in the information captured by certain measurements. Such multicollinearity can impact model interpretability and may influence feature selection decisions.

Understanding these inter-feature relationships is important for guiding downstream modeling choices. In particular, the presence of correlated inputs supports the use of tree-based ensemble models, which are generally robust to multicollinearity and can effectively leverage redundant signals without requiring explicit feature elimination.

This analysis also provides a foundation for potential dimensionality reduction or feature pruning in later stages of the project, if required.
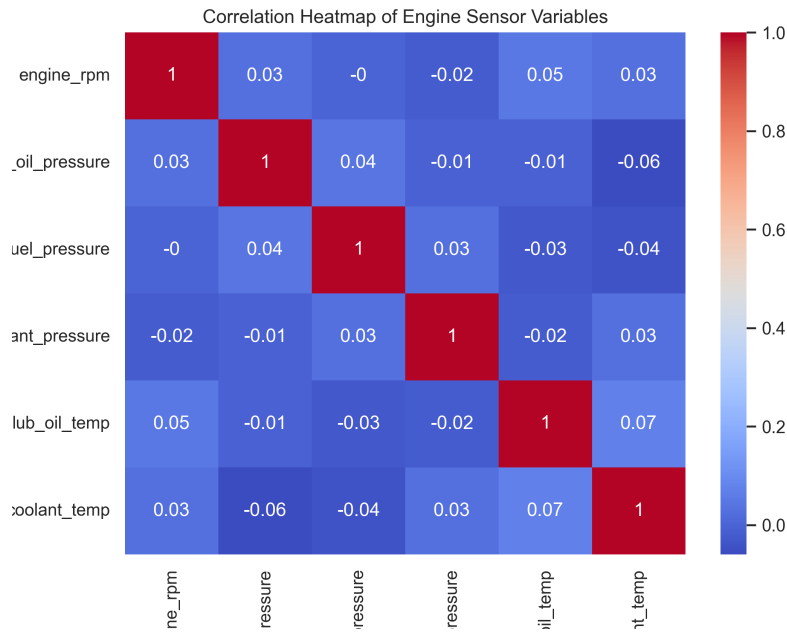
Figure 5: Correlation heatmap of engine sensor variables

3.6 EDA Insights and Observations

- Engine failure events are relatively rare, indicating class imbalance and the need for recall-focused evaluation metrics.

- Pressure- and temperature-related sensors show systematic shifts prior to failure, highlighting their predictive relevance.

- Several sensor variables are strongly correlated, suggesting redundancy but supporting the use of tree-based ensemble models.

- The observed non-linear relationships motivate the use of ensemble learning approaches over linear models.

These insights directly inform the data preparation strategy and guide the selection of appropriate modeling techniques in the subsequent stages of the project.

4   Data Preparation

This section describes the steps undertaken to prepare the dataset for machine learning modeling. The objective of data preparation is to ensure data quality, remove non-informative attributes, and create reproducible training and testing datasets suitable for experimentation and evaluation.

4.1   Data Loading

The dataset was loaded directly from the Hugging Face Dataset Hub to ensure consistency with the registered data source. Loading the data from a centralized repository guarantees reproducibility and ensures that all experiments are conducted on a version-controlled dataset rather than local copies.

4.2   Data Cleaning and Feature Selection

An initial data cleaning step was performed to remove columns that are not relevant for predictive modeling. Since the objective is to predict engine failure based on sensor behavior, only numerical sensor variables and the target variable were retained.

   The dataset did not contain missing values across the selected features, eliminating the need for imputation. No feature scaling or normalization was applied at this stage, as the subsequent modeling phase focuses on tree-based algorithms that are inherently robust to differences in feature scale and distribution.

4.3   Train–Test Split

The cleaned dataset was split into training and testing subsets to enable unbiased model evaluation. A stratified split strategy was employed to preserve the original class distribution of the target variable in both subsets, which is particularly important given the class imbalance observed during exploratory data analysis.

   The training and testing datasets were saved locally to maintain a clear separation between raw, processed, and modeling-ready data.

4.4   Dataset Versioning and Upload

To maintain reproducibility and support experiment tracking, the processed training and testing datasets were uploaded back to the Hugging Face Dataset Hub. This ensures that all modeling experiments reference a fixed and verifiable version of the prepared data.

   By registering both raw and processed datasets on the Hugging Face platform, the data preparation workflow remains transparent, auditable, and reproducible.

Methodology Summary

The data preparation process involved loading version-controlled datasets from the Hugging Face platform, selecting relevant sensor variables, and performing a stratified train–test split to preserve class distribution. No feature scaling was applied, as tree-based models were selected for downstream modeling.

5 Model Building with Experimentation Tracking

This section describes the development, tuning, evaluation, and registration of machine learning models for predicting engine failure. The objective is to establish a transparent and reproducible experimentation workflow that enables systematic comparison of model performance.

5.1 Data Loading

The training and testing datasets were loaded directly from the Hugging Face Dataset Hub to ensure consistency with the versioned outputs of the data preparation stage. This approach guarantees that all modeling experiments are conducted on fixed and auditable datasets.

5.2 Model Selection and Baseline

Based on insights from exploratory data analysis, tree-based machine learning algorithms were selected due to their ability to model non-linear relationships, robustness to outliers, and tolerance to correlated features. A Decision Tree classifier was used as a baseline model to establish a performance reference point.

5.3 Hyperparameter Tuning and Experiment Tracking

To improve predictive performance, ensemble-based models were trained with systematic hyperparameter tuning. Key model parameters were varied and logged to enable structured experimentation and comparison across model configurations.

All hyperparameters and evaluation metrics were explicitly recorded to ensure transparency and reproducibility of the experimentation process.

5.4 Model Evaluation

Model performance was evaluated on the held-out test dataset using metrics appropriate for imbalanced classification problems. Precision, recall, F1-score, and ROC-AUC were used to assess the trade-off between failure detection and false maintenance alerts.

The confusion matrix was analyzed to understand error distribution and to quantify the cost of missed failure events.

Table 1: Comparison of baseline and final model performance on test data

Model	Recall	F1-score	ROC-AUC
Decision Tree	0.678441	0.675971	0.558750
Random Forest	0.900122	0.770059	0.699443

5.5 Best Model Selection and Registration

The best-performing model was selected based on recall and F1-score, reflecting the business priority of minimizing undetected engine failures. This model was subsequently registered on the Hugging Face Model Hub, along with its configuration and performance metrics.

Registering the model on the Hugging Face platform ensures reproducibility, enables version control, and supports future deployment and evaluation workflows.
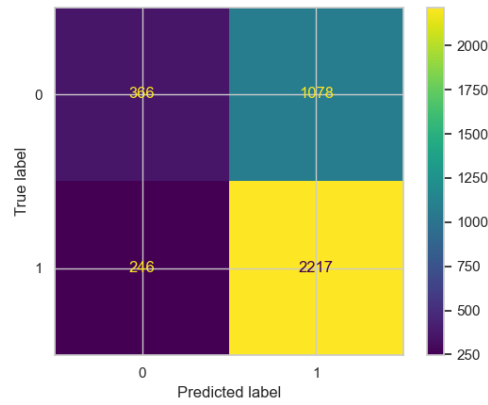
Figure 6: Confusion matrix for the selected Random Forest model on the test dataset

Methodology Summary

Tree-based classification models were selected due to their robustness to non-linear relationships, outliers, and correlated features. Hyperparameter tuning was performed using grid search with recall as the primary optimization metric, reflecting the business priority of minimizing missed failure events.

6　Interim Findings and Next Steps

The interim analysis demonstrates that engine failure events can be predicted with reasonable accuracy using sensor-based machine learning models. Exploratory analysis confirmed the presence of informative sensor signals related to pressure and temperature variations prior to failure events. The selected Random Forest model achieved improved recall compared to baseline models, indicating its effectiveness in capturing failure patterns.

In the next phase of the project, further enhancements will include model deployment, automated experiment workflows, and extended evaluation on unseen data. Additional analysis will also focus on translating model outputs into actionable maintenance recommendations and operational insights.

A　Appendix

This appendix contains supplementary material supporting the analyses presented in the main report. All raw code, extended outputs, and detailed experiment logs are maintained in the accompanying executed Python notebook submitted in HTML format.