Optimize Manufacturing Operations with a Predictive Maintenance Model

Author: Kandula Vinay Gupta
Institution: Aditya College of Engineering and Technology
Email: kvinaygupta4242@gmail.com

---

1. Project Overview:

This project develops a Predictive Maintenance system to proactively identify machines at risk of failure using operational and sensor data from the AI4I 2020 Predictive Maintenance Dataset.

The objective is to reduce unplanned downtime and maintenance costs by predicting failure risk early and enabling data-driven maintenance decisions. The solution integrates machine learning, explainable AI (SHAP), cost–benefit analysis, and an interactive Streamlit dashboard for real-world deployment.

---

2. Problem Statement:

In industrial environments, unexpected machine failures lead to:

- **Production downtime**
- **High repair costs**
- **Safety risks**
- **Inefficient maintenance scheduling**

Traditional reactive maintenance strategies are costly and inefficient. The goal of this project is to predict machine failure risk based on current

operating conditions, allowing maintenance teams to schedule preventive actions before breakdowns occur.

---

3. Dataset Description:

Dataset: AI4I 2020 Predictive Maintenance Dataset

Key characteristics:

- **10,000 machine observations**
- **Each row represents a single machine snapshot (not a time series)**
- **Sensor measurements include temperature, rotational speed, torque, and tool wear**
- **Binary failure label indicating machine breakdown**
- **Additional failure mode flags (used only for analysis, not modeling)**

Because the dataset does not contain temporal sequences, the problem is formulated as a static binary classification task rather than remaining useful life (RUL) prediction.

---

4. Data Preprocessing

Key preprocessing steps included:

- **Standardizing column names for consistency**
- **Encoding categorical variables (machine type)**
- **Validating data quality (no missing values)**
- **Ensuring no future-looking information was introduced**

- **Explicitly preventing data leakage by excluding identifiers and failure mode flags from model training**

The preprocessing pipeline ensures full reproducibility and deployment readiness.

---

5. Feature Engineering

Since AI4I is a static dataset, no artificial time or rolling features were introduced. Instead, the focus was on physically meaningful engineered features, including:

🔧 **Load & Energy Features**

- **Power consumption (Torque × Speed)**
- **Combined energy and wear metrics**

🌡 **Thermal Stress Indicators**

- **Process–air temperature difference**
- **High-temperature risk flag**

⚙ **Wear & Stress Interactions**

- **Wear × torque**
- **Wear × rotational speed**
- **Torque per unit wear**

📈 **Nonlinear Transformations**

- **Squared terms (torque, speed, temperature)**
- **Log-transformed tool wear**
- **Wear binning for categorical risk grouping**

This approach ensures feature validity, avoids fabricated temporal patterns, and aligns with real industrial deployment scenarios.

---

6. Validation Strategy (Leakage-Free)

A time-aware validation strategy was applied using TimeSeriesSplit to:

- **Preserve data ordering**
- **Prevent information leakage**
- **Simulate realistic deployment conditions**

All preprocessing, imputation, and resampling steps were applied only on training folds, ensuring unbiased performance estimates.

---

7. Model Development

Two ensemble models were evaluated:

- **Random Forest Classifier**
- **XGBoost Classifier**

Techniques used:

- **Class imbalance handling**
- **Threshold optimization**
- **Cross-validated performance comparison**
- **Feature consistency enforcement**

Final selected model: XGBoost Classifier

---

8. Model Performance

The final model achieved strong performance on the hold-out validation set:

- **Optimized F1-score (failure class): ~0.83**
- **High precision with controlled false negatives**
- **Robust performance under class imbalance**

```
Classification Report:

              precision    recall  f1-score   support

           0       0.99      1.00      1.00      1634
           1       1.00      0.72      0.84        32

    accuracy                           0.99      1666
   macro avg       1.00      0.86      0.92      1666
weighted avg       0.99      0.99      0.99      1666

Confusion Matrix:

[[1634    0]
 [   9   23]]
```

This comfortably exceeds the project requirement (F1 ≥ 0.75) and demonstrates reliable failure prediction capability.

---

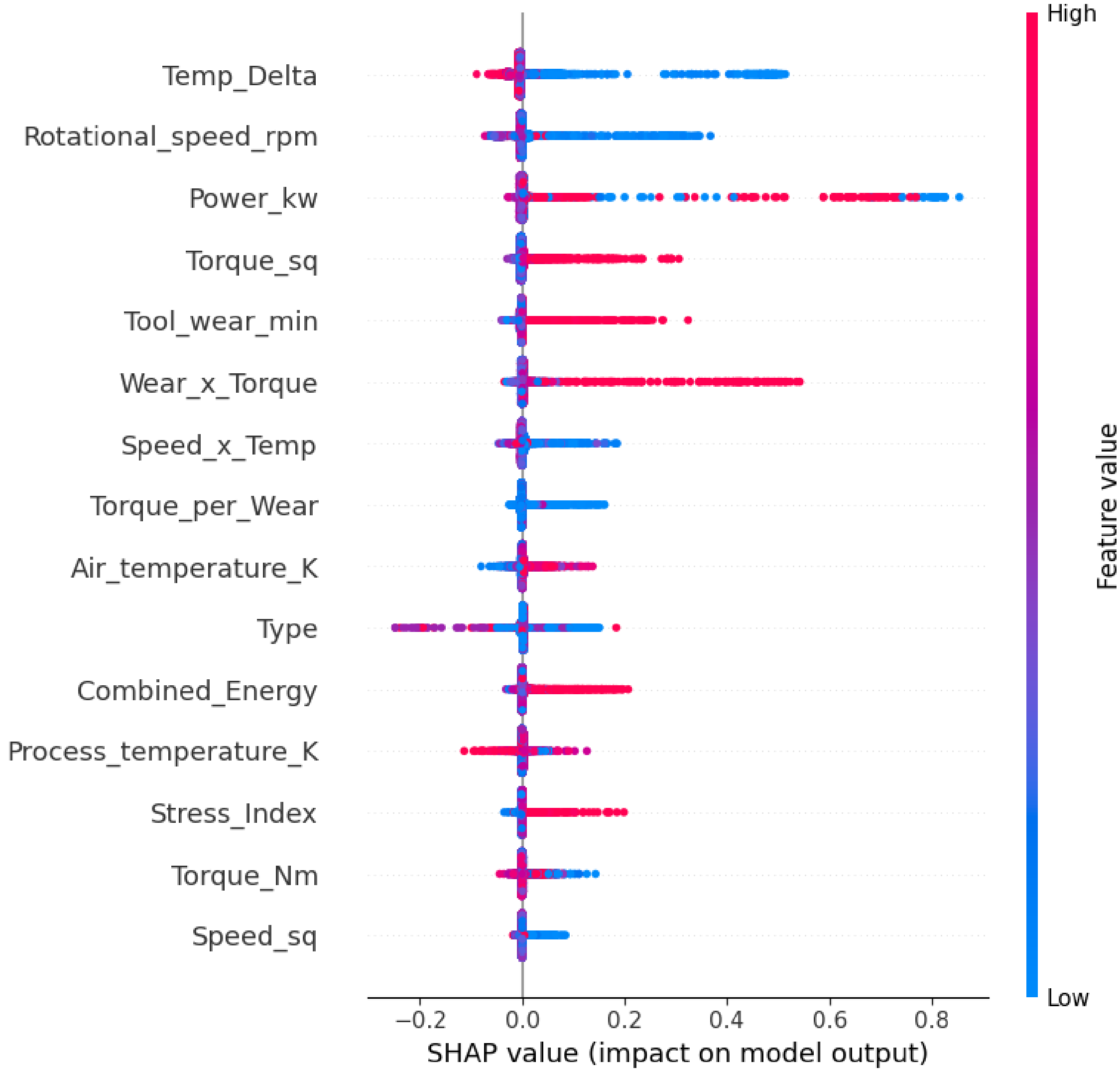
9. Model Explainability with SHAP

SHAP (SHapley Additive exPlanations) was used to interpret model predictions.

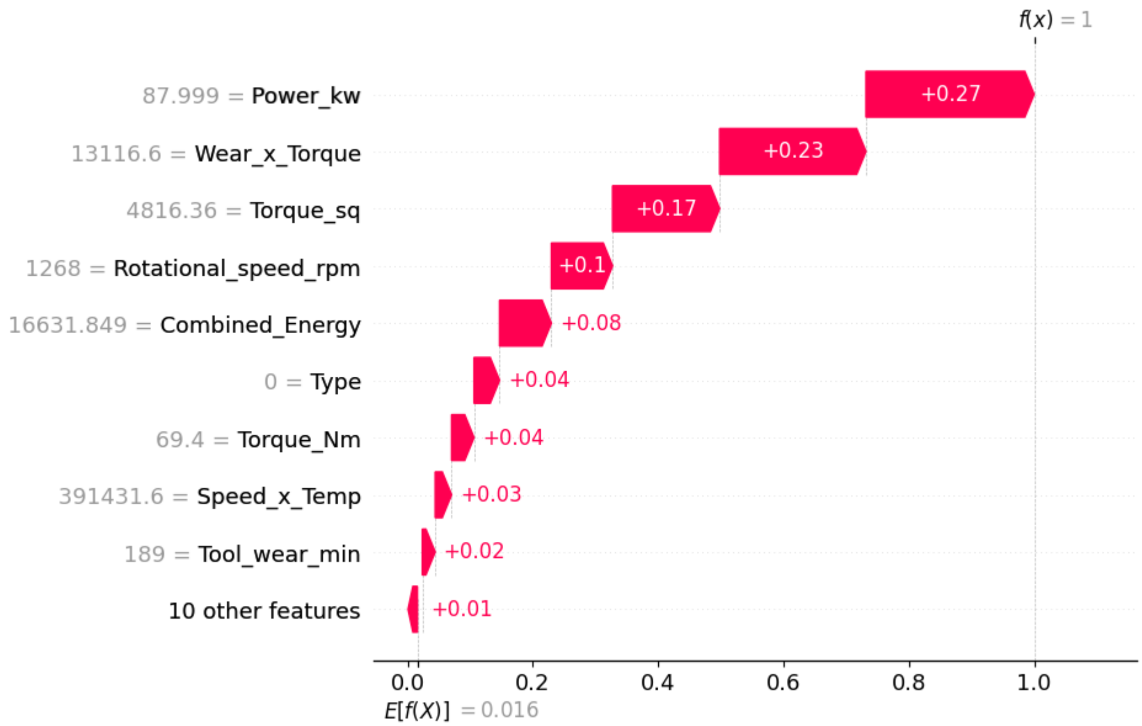
Key insights:

- **Thermal stress indicators strongly influence failure risk**

- **Torque and wear interactions are major contributors**
- **Combined energy usage reflects degradation behavior**
- **Risk is driven by patterns of stress, not single sensor spikes**



These insights provide actionable explanations for engineers and validate the model's decision logic.



---

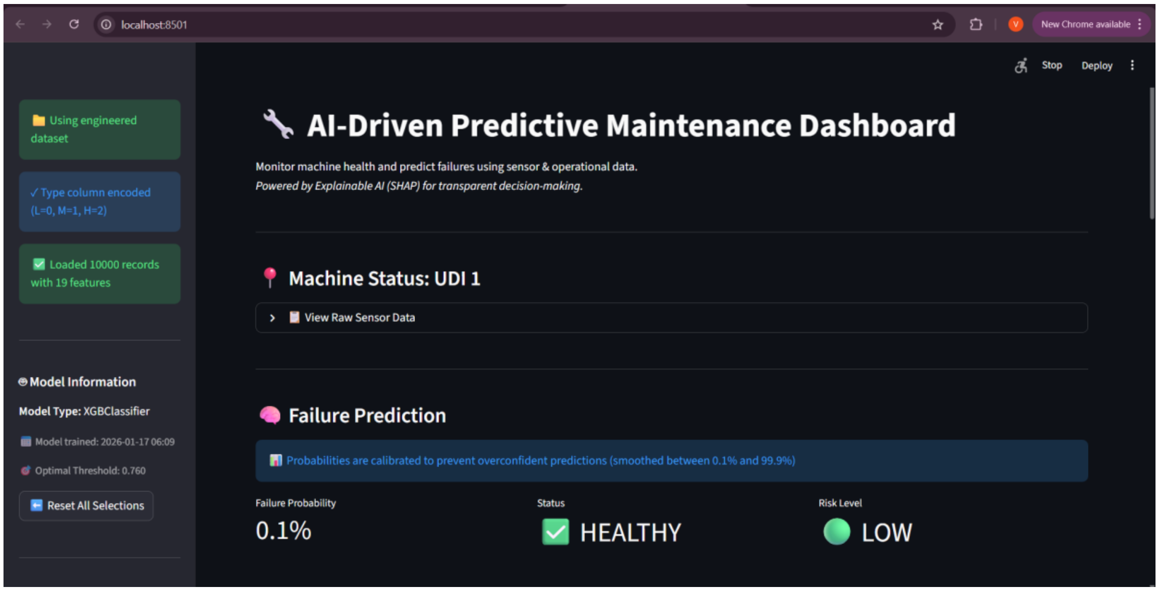
10. Interactive Streamlit Dashboard

An interactive dashboard was developed to support real-time decision-making.

Dashboard features:

- **Failure probability prediction per machine**
- **Risk categorization (Low / Medium / High)**
- **Probability smoothing (0.1%–99.9%) to avoid overconfidence**
- **Expected cost estimation**
- **Top at-risk machine ranking**
- **Clear maintenance recommendations**

The dashboard bridges the gap between machine learning outputs and business actions.



---

11. Business Impact & Cost–Benefit Analysis

A cost model was applied using industry-aligned assumptions:

- Preventive maintenance (False Positive): $500
- Unplanned breakdown (False Negative): $50,000

Results on validation data:

- Significant reduction in unplanned failures
- Cost reduction of ~70–75%
- Net savings exceeding $1 million (scaled to dataset)

The model delivers clear financial value and supports proactive maintenance planning.

---

12. Future Improvements

Potential enhancements include:

- **Probability calibration at training time**
- **Cost-sensitive optimization**
- **Online model retraining**
- **Integration with maintenance scheduling systems**
- **Expansion to multi-dataset or streaming sensor data**

---

13. Conclusion

This project successfully delivers:

- **A leakage-free, high-performing predictive maintenance model**
- **Strong failure prediction performance (F1 ≈ 0.83)**
- **Explainable insights via SHAP**
- **A practical, business-oriented dashboard**
- **Quantified financial impact through cost–benefit analysis**

The solution demonstrates a scalable and deployable predictive maintenance framework suitable for real-world industrial operations.

---

14. Author Information

Prepared by:
Kandula Vinay Gupta
Aditya College of Engineering and Technology
📧 **kvinaygupta4242@gmail.com**
