FSE 570 Data Science Capstone Project Proposal

Data-Driven Anomaly Detection and Risk-Aware Maintenance Scheduling for Smart Industrial Systems

Team Members: Anoushka Jaydas Dighe, Deva Siva Kanth Tavvala, Mohit Kumar Petla, Umang Rajnikant Bid, Urvansh Jignesh Shah

Course: FSE 570

Problem Statement

Modern manufacturing facilities face critical operational challenges from unexpected equipment failures that result in costly production downtime (15-20 hours per incident), maintenance expenses ($50K-$200K per unplanned outage), safety hazards, and inefficient resource allocation. Despite continuous high-frequency sensor data collection from industrial equipment, most monitoring systems operate reactively using threshold-based alarms that generate excessive false positives, or employ predictive models that forecast failures without providing actionable maintenance guidance under operational constraints.

Core Challenge: How can industrial systems detect abnormal equipment behavior early, quantify failure risk with uncertainty estimates, and optimally decide which machines to maintain and when, given limited maintenance resources?

Value Proposition: This project delivers an end-to-end decision support system that transforms maintenance operations from reactive to proactive. The system provides manufacturing operators with anomaly alerts, calibrated failure probabilities, and data-driven maintenance schedules that minimize total operational cost (downtime + maintenance expenses) while maximizing equipment availability under resource constraints (e.g., 3 maintenance crews for 50+ machines).

Industry Impact: Manufacturers implementing predictive maintenance report 25-30% maintenance cost reduction, 70% decrease in unplanned downtime, and 10-15% equipment lifespan extension. This project addresses the critical gap between failure prediction and actionable decisions—enabling strategic resource allocation in multi-machine manufacturing environments.

Data Sources

Dataset 1: High-Frequency Sensor Telemetry
Source: NASA C-MAPSS Jet Engine Dataset + Kaggle Industrial Equipment Monitoring
Content: 21+ sensors (temperature, pressure, vibration, speed, power) across 100+ units with 260+ cycles each (32,000+ observations)
Challenges: High dimensionality, missing values, noise filtering, temporal alignment

Dataset 2: Maintenance and Failure Logs
Source: C-MAPSS labels + synthetic maintenance records
Content: Remaining Useful Life labels, repair history, downtime, costs, failure types (bearing fault, imbalance, overheating); 100+ failures, 1,600+ logs
Challenges: Censored observations, imbalanced classes, temporal integration

Dataset 3: Operational Context
Content: Machine specs, production schedules, operational settings, sensor mapping, crew availability
Challenges: Heterogeneous data types, spatial-temporal alignment

Integration: Temporal windowing (30-cycle sequences), feature engineering (statistics, trends), timestamp-based joins, RUL calculation, temporal train-test split (70-15-15).

Methodology

Our approach integrates advanced machine learning, neural networks, Bayesian modeling, and decision optimization:

1. Anomaly Detection (Unsupervised Deep Learning)
Technique: LSTM-based temporal autoencoders learn compressed representations of normal sensor patterns; anomaly score = reconstruction error
Justification: Captures complex non-linear temporal patterns without labeled anomalies; addresses cold-start problem

2. Failure Risk Prediction with Uncertainty (Supervised + Bayesian)
Models:

- LSTM sequence models: Predict failure probability within h cycles using 30-cycle windows
- XGBoost regression: Remaining Useful Life estimation with engineered features
- Bayesian Weibull survival analysis: Time-to-failure with uncertainty quantification

Validation: Walk-forward time-series cross-validation; metrics include RMSE (RUL), F1-score (classification), calibration curves; 90%/95% credible intervals from Bayesian posteriors
Justification: LSTMs capture long-range dependencies; Bayesian methods provide calibrated uncertainty for risk-aware decisions

3. Explainability
Methods: SHAP values for feature attribution; attention visualization for temporal importance; sensor ranking
Justification: Engineering validation requires interpretable insights for model trust and root cause diagnosis

4. Maintenance Decision Optimization (MILP)
Formulation: Minimize expected cost = Σ(failure_risk × downtime_cost + maintenance_cost)
Constraints: Crew capacity (≤3 concurrent jobs), time windows, safety thresholds
Solver: Python PuLP/Google OR-Tools
Justification: Converts predictions to actionable schedules under operational constraints; MILP provides optimal solutions

Workflow: Data preprocessing → Model training (autoencoder, LSTM, XGBoost, Bayesian) → Real-time inference (anomaly scores, failure probabilities) → MILP optimization → Maintenance schedules → Evaluation via simulation