Fault Detection and Predictive Maintenance of Industrial Gas Turbine Engines

Garey Salinas
garey.salinas@colorado.edu
University of Colorado-Boulder
Boulder, Colorado, USA

Abstract

This project aims to use sensor data to develop a machine learning-based fault detection system for industrial gas turbine engines. By leveraging a publicly available dataset from Kaggle, which includes labeled operational and faulty states, we will apply classification models, outlier detection, and time-series segmentation techniques to identify early signs of compressor failure. The system, by reducing downtime, optimizing maintenance schedules, and improving efficiency, promises to significantly enhance the operational outcomes in midstream pipeline operations where gas turbines are essential for compression.

Keywords

fault detection, gas turbines, predictive maintenance, machine learning

1 Introduction

Industrial gas turbine engines are crucial in midstream pipeline operations, powering gas compressors. Unplanned failures can lead to significant operational disruptions, economic losses, and safety hazards. This project, however, takes a proactive approach. It explores sensor-based predictive maintenance using machine learning to detect faults before they escalate into serious issues. The use of data mining for predictive diagnostics not only enhances reliability but also reduces maintenance costs and downtime.

2 Literature Review

According to Lei et al., using sensor data for predictive maintenance and fault detection has attracted significant interest in the industrial domain. Machine learning models such as Support Vector Machines (SVM), Random Forests (RF), or deep learning models [4] are currently widely applied in machine learning processes for the diagnostics of rotating machinery. We can get good results even with basic time-series data from the sensors.

Studies such as Babu et al. (2016) present the importance of feature engineering, particularly time-domain and frequency-domain features extracted from temperature and vibration signals, as applicable for condition monitoring. For example, pressure ratios and temperature differentials have been helpful for early failure detection.

Time-series segmentation is another area of active research. Malhotra et al. (2016) utilized Long Short-Term Memory (LSTM) networks to capture temporal patterns. However, simpler statistical methods like rolling averages and trend decomposition also yield interpretable results for fault trend analysis.

Outlier detection remains essential in predictive maintenance. Approaches such as Isolation Forest (Liu et al., 2008) and DBSCAN are robust to noisy data and have been effectively used in equipment anomaly detection, especially where rare fault labels make supervised learning difficult. These methods are particularly valuable in identifying collective and contextual outliers from sensor streams.

This project aims to integrate these studies' insights and apply supervised and unsupervised learning to industrial gas turbine health monitoring.

3 Proposed Methodology

The dataset used for this project is the Kaggle Gas Turbine Engine Fault Detection Dataset (https://www.kaggle.com/datasets/ziya07/gas-turbine-engine-fault-detection-dataset). The dataset includes actual sensor data (e.g.,temperatures, pressures, rotational speeds, etc.) from gas turbine engines and labeled fault types. This makes it an excellent fit for supervised machine learning.

I aim to build a classification model to detect failures in gas compressors. This applies to midstream pipeline operations, where gas turbines are typically used as a compressor driving mechanism.

Dataset

- **Source:** Kaggle Gas Turbine Engine Fault Detection Dataset
- https://www.kaggle.com/datasets/ziya07/gas-turbine-engine-fault-detection-dataset

Data Preprocessing

- Missing value handling
- Sensor reading normalization
- Feature engineering (pressure ratios, temperature deltas, rolling averages, etc.)

Modeling Techniques

- **Supervised Models:** Random Forest, XGBoost, Support Vector Machines
- **Unsupervised Models:** Isolation Forest, DBSCAN
- **Time-Series Segmentation:** Apply rolling window statistics and operational cycle segmentation to detect temporal degradation patterns

Tools & Libraries

- Python (Pandas, Scikit-learn, XGBoost, Seaborn, Plotly)
- Jupyter Notebook
- Optional: Streamlit for interactive results dashboard

4 Evaluation Plan

- **Classification Metrics**

- Accuracy
- Precision
- Recall
- F1-Score

- **Anomaly Detection Metrics**
  - ROC-AUC
  - Confusion Matrix
  - Visualization of Anomalies
- **Validation Strategy**
  - K-Fold Cross-Validation to ensure robust model performance across different subsets of the data
- **Interpretability**
  - Use SHAP values or feature importances to explain and interpret model predictions

5 Timeline

This timeline outlines the key phases of the project, beginning with literature review and data preparation, followed by model development and anomaly detection. The final days are allocated to compiling results and preparing the final report and presentation. Table 1 summarizes the major project milestones. The project is scheduled for completion by April 28, 2025, aligning with the course submission deadline.

6 Citations and Bibliographies

This project builds on a range of research studies and industry data sources. The core background for this work includes a comprehensive survey on anomaly detection by Chandola et al. [2] and a foundational review of machinery health prognostics by Lei et al. [5]. Isolation Forest, introduced by Liu et al. [6], is among the primary algorithms employed in our anomaly detection phase. Time-series segmentation is informed by Malhotra et al.'s work on TimeNet [7], while the importance of predictive maintenance is reinforced by recent market insights from IoT Analytics [4]. Additionally, an Applied Sciences journal article [1] presents a helpful overview of current machine learning models in gas turbine anomaly detection. The dataset used for experimentation is the Gas Turbine Engine Fault Detection Dataset hosted on Kaggle [3].

References

[1] Various Authors. 2024. Review of Machine Learning Models for Gas Turbine Anomaly Detection. *Applied Sciences* 14, 11 (2024), 4551. doi:10.3390/app14114551

[2] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly Detection: A Survey. *ACM Computing Surveys (CSUR)* 41, 3 (2009), 1–58. doi:10.1145/1541880.1541882

[3] Ziyao Chen. 2020. Gas Turbine Engine Fault Detection Dataset. https://www.kaggle.com/datasets/ziya07/gas-turbine-engine-fault-detection-dataset. Accessed: 2025-04-16.

[4] IoT Analytics. 2023. Predictive Maintenance Market Report 2023. https://iot-analytics.com/predictive-maintenance-market. Accessed: 2025-04-16.

[5] Yaguo Lei, Naipeng Li, Lin Guo, Ning Li, Tao Yan, and Jing Lin. 2020. Machinery Health Prognostics: A Systematic Review from Data Acquisition to RUL Prediction. *Mechanical Systems and Signal Processing* 138 (2020), 106761. doi:10.1016/j.ymssp.2019.106761

[6] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation Forest. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*. IEEE, 413–422. doi:10.1109/ICDM.2008.17

[7] Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. 2017. TimeNet: Pre-trained Deep Recurrent Neural Network for Time Series Classification. *arXiv preprint arXiv:1706.08838* (2017). https://arxiv.org/abs/1706.08838

Table 1: Proposed Project Timeline

Phase	Dates	Tasks
Project Planning & Literature Review	Apr 16 – Apr 18	Finalize proposal, conduct literature review, outline methods
Data Preparation & EDA	Apr 18 – Apr 20	Clean dataset, perform exploratory data analysis, and engineer features
Model Building (Classification)	Apr 20 – Apr 23	Train and evaluate Random Forest, XGBoost, and SVM models
Outlier Detection & Time-Series Analysis	Apr 23 – Apr 25	Apply Isolation Forest, DBSCAN, and segment time-series data for degradation pattern analysis
Final Report & Presentation	Apr 25 – Apr 28	Compile results, write final report, and prepare presentation materials