Detecting Anomalous Activity in a Ship’s Engine

Table of Contents

1. Introduction

Shipping companies rely on the seamless performance of their vessels to ensure operational efficiency, timely deliveries, and cost-effectiveness. Undetected engine malfunctions can lead to severe breakdowns, increased maintenance costs, and potential safety hazards. This report outlines a data-driven approach to identifying anomalies in ship engine performance, allowing for proactive maintenance and reducing unplanned downtime.

2. Business Context

Engine failures disrupt shipping schedules, incur high repair costs, and impact customer satisfaction. Early detection of anomalies can prevent such issues, ensuring business continuity and profitability. This project evaluates a combination of anomaly detection methods to identify abnormal engine behavior before it results in costly repairs or service interruptions.

3. Approach and Methodology

This project follows a structured data-driven approach:

- **Data Collection**: Analysis of six critical engine performance indicators.
- **Exploratory Data Analysis (EDA)**: Assessing data quality and identifying key trends.
- **Feature Engineering**: Standardisation of data for consistency in analysis.
- **Anomaly Detection Methods**: Testing statistical and machine learning approaches to detect deviations in engine behavior.
- **Model Evaluation**: Selecting the most reliable approach based on anomaly detection accuracy and business relevance.

4. Data Analysis and Feature Exploration

The dataset underwent an initial assessment to ensure completeness and reliability:

- No missing or duplicate values were detected.
- Key engine metrics were analysed to understand normal operating conditions and potential problem areas.
- Outlier detection indicated that lubrication oil temperature and fuel pressure exhibited the most frequent anomalies, making them critical monitoring parameters.

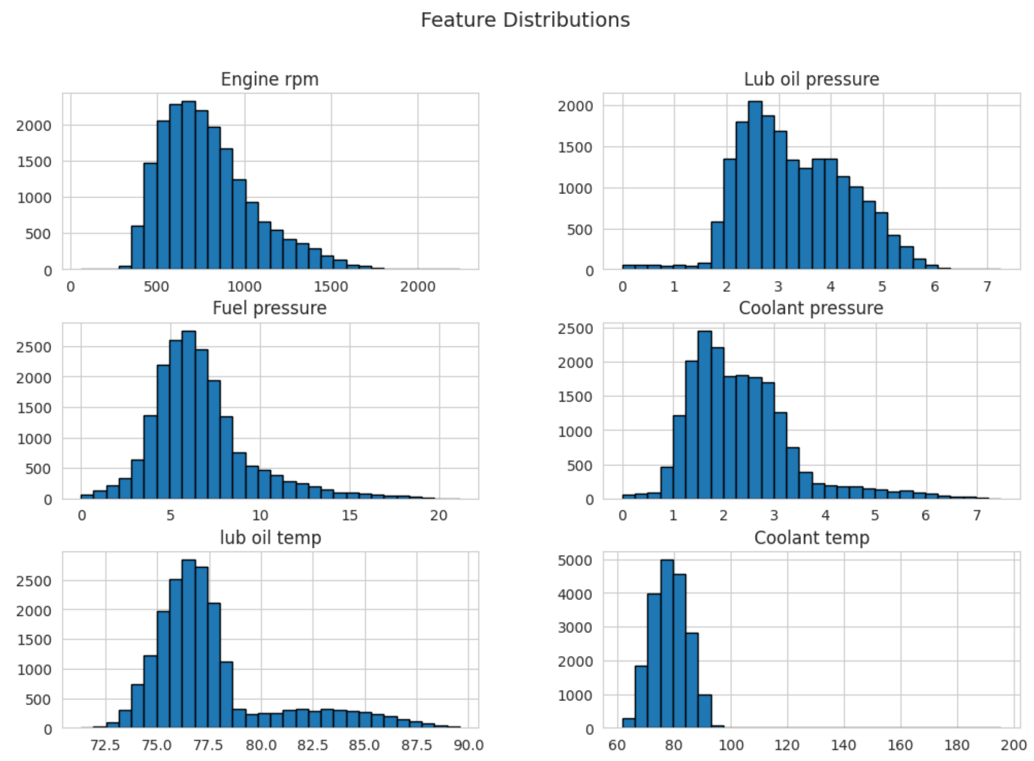

Figure 1: Feature Distributions

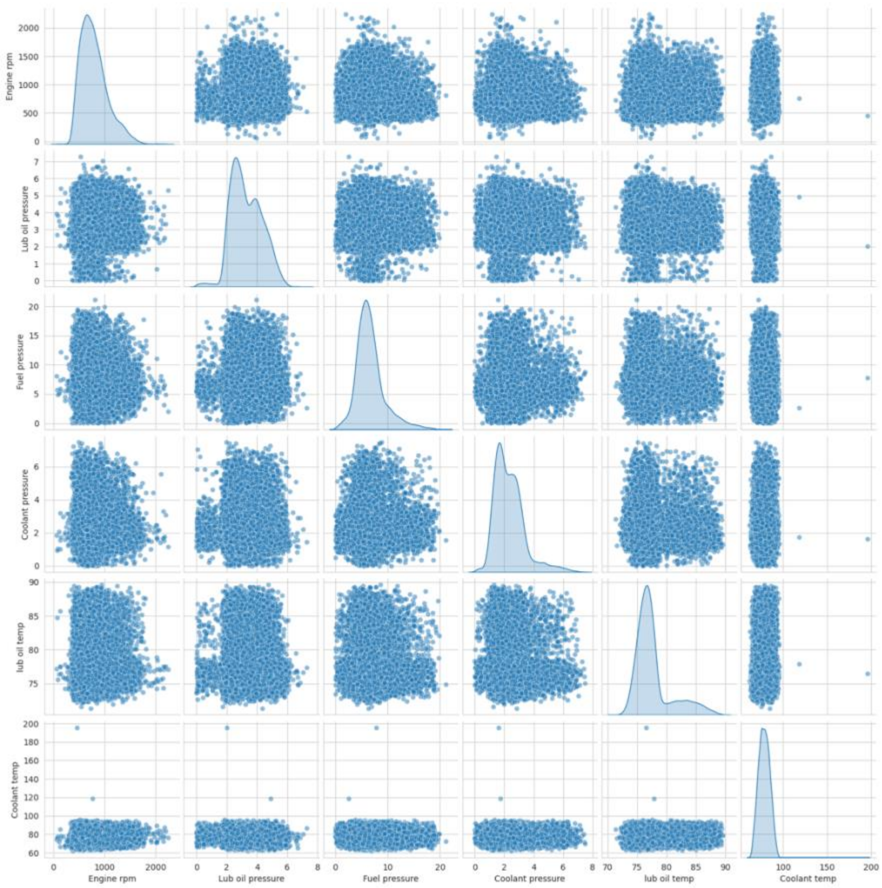

Figure 2: Pairwise Scatterplots

5. Anomaly Detection Techniques

5.1 Statistical Approach: Interquartile Range (IQR)

A straightforward statistical method, the IQR technique identifies extreme values that fall significantly outside expected operational ranges. The method detected anomalies in 2.16% of the data, with fuel pressure and lubrication oil temperature being the most affected variables.

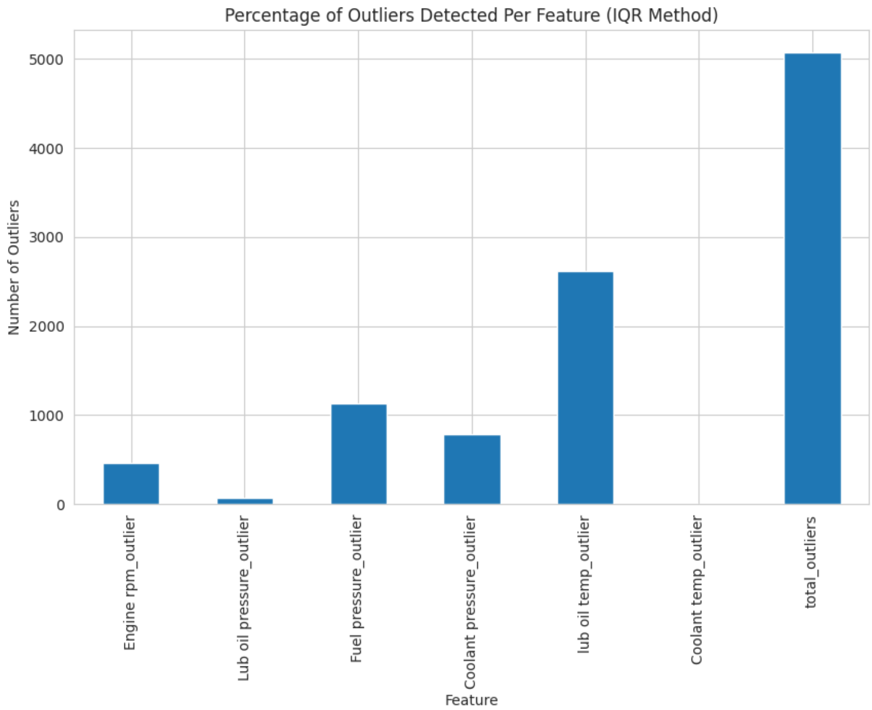

Figure 3: Percentage of Outliers Detected Per Feature (IQR Method)

5.2 Machine Learning Approaches

One-Class SVM

This model was fine-tuned to balance anomaly detection with minimizing false positives. The best configuration identified 5% of the dataset as anomalies, aligning with expectations for early fault detection.

nu	gamma	num_anomalies	anomaly_rate
0.01	scale	198	1.013565
0.01	auto	198	1.013565
0.05	scale	976	4.996161
0.05	auto	976	4.996161
0.10	scale	1954	10.002560
0.10	auto	1954	10.002560

Figure 4: One-Class SVM Hyperparameter Tuning Results

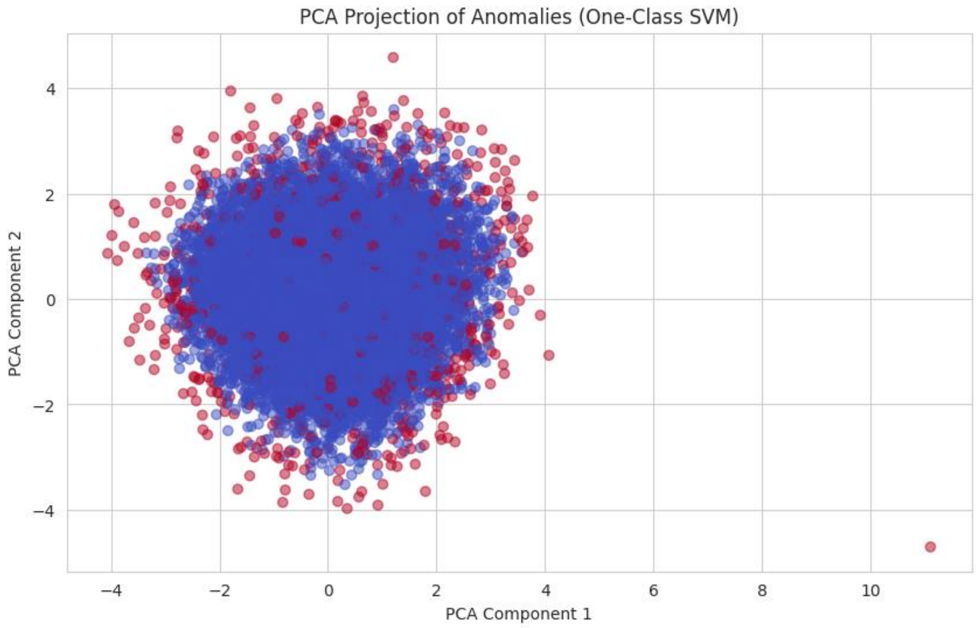

Figure 5: PCA Projection of Anomalies (One-Class SVM)

Isolation Forest

Isolation Forest was tested at different sensitivity levels, with a contamination threshold of 3% providing a balanced and reliable detection rate.

contamination	num_anomalies	Anomaly_rate
0.01	196	1.003327
0.03	587	3.004863
0.05	977	5.001280

Figure 6: Isolation Forest Hyperparameter Tuning Results

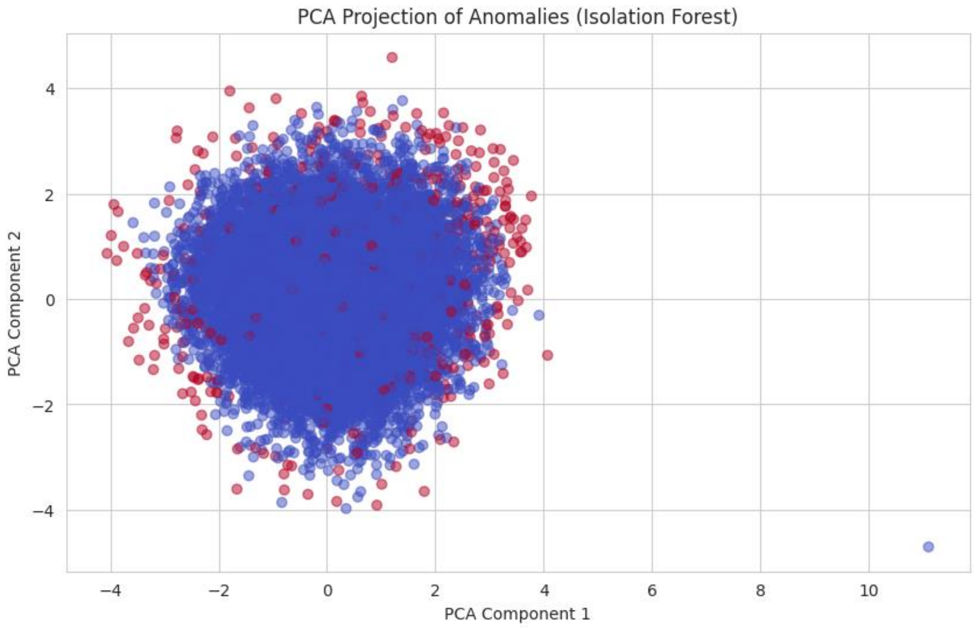

Figure 7: PCA Projection of Anomalies (Isolation Forest)

6. Insights and Findings

- **Lubrication oil temperature and fuel pressure** are the most critical factors for anomaly detection.
- Statistical methods (IQR) provide a simple but rigid approach to anomaly detection.
- Machine learning models (One-Class SVM and Isolation Forest) enable a more flexible and robust analysis.
- **Isolation Forest** emerged as the most effective model, balancing precision and reliability in detecting anomalies.

7. Best Performing Model Selection

Among the tested methods:

- **One-Class SVM** detected 5% anomalies, capturing more potential failures but with higher false positive risk.
- **Isolation Forest** detected 3% anomalies, providing a more targeted and actionable set of alerts.
- **IQR**, while simple, lacked adaptability and was less effective in identifying complex patterns.

The **Isolation Forest model** was selected as the best approach due to its reliability, adaptability, and business relevance.

8. Business Implications and Recommendations

Early detection of anomalies in ship engines translates directly into operational efficiencies and cost savings. By integrating machine learning-driven anomaly detection into existing maintenance workflows, businesses can:

- **Reduce Unplanned Downtime**: Predictive maintenance allows for scheduled interventions, avoiding costly engine failures during operations.
- **Optimize Fuel Efficiency**: Addressing early warning signs of engine inefficiencies can lead to reduced fuel consumption.
- **Lower Maintenance Costs**: Identifying anomalies early minimizes the need for extensive repairs, reducing overall maintenance expenditure.
- **Enhance Safety**: Preventing major engine malfunctions ensures the safety of crew members and cargo.

To maximize the impact of this system, businesses should consider:

- **Deploying real-time anomaly detection tools** that integrate with onboard monitoring systems.

- **Setting up alert thresholds** based on historical anomaly patterns to refine predictive accuracy.
- **Training maintenance teams** to interpret model outputs and act on them efficiently.

9. Evaluating 2D Visualisations

PCA visualisations provided clear insights into detected anomalies, helping to separate normal engine behavior from potential failures. **Isolation Forest's PCA projection** proved particularly effective in visually distinguishing anomalous points from normal operations, making it a valuable tool for decision-makers.

10. Conclusion

The implementation of an anomaly detection system in ship engines offers significant benefits for predictive maintenance. By leveraging **Isolation Forest**, shipping companies can identify early warning signs of engine malfunctions, reducing operational risks and maintenance costs. Future improvements could focus on integrating real-time monitoring systems and refining detection thresholds based on industry-specific requirements.