DETECTING THE ANOMALOUS ACTIVITY OF A SHIP'S ENGINE

Report

DATE :
March 2025

Table Of Content

Introduction

Efficient ship engine performance is vital, as failures can cause costly delays, safety risks, and higher expenses. Anomalies in rpm, oil pressure, or coolant temperature may signal malfunctions that, if undetected, lead to severe failures. Early detection through data analysis helps reduce maintenance costs, improve fuel efficiency, and ensure timely deliveries, enhancing profitability and trust.

This report develops an anomaly detection system by analyzing key engine metrics like rpm, fuel pressure, and oil temperature. It evaluates statistical and machine learning methods to identify the most effective approach. While focusing on anomaly detection and business impact, external factors like environmental influences are beyond its scope.

Problem Statement

Neglected ship engine maintenance results in undetected anomalies, leading to inefficiencies, safety hazards, and higher costs from unexpected downtime and fuel waste. To address this, the company seeks to implement a machine learning-based anomaly detection system to predict failures early, minimize maintenance expenses, and improve delivery reliability, strengthening stakeholder confidence.

Methodology

To detect anomalous activity in a ship's engine, a data-driven approach was employed, leveraging machine learning techniques and statistical analysis. Meaningful insights, including the means, median and 95th percentile was used to, to capture fluctuations in engine performance.

Several anomaly detection techniques were explored, including statistical methods like IQR and machine learning algorithms such as Isolation Forest and one-class SVM. The models were trained using historical data to establish a baseline of normal engine behavior. Finally, visualizations such as PCA plots, histograms and boxplots are were used to interpret the results and provide actionable insights.

Data Statistics

The dataset tracks six key engine metrics: RPM, Lubrication Oil Pressure, Fuel Pressure, Coolant Pressure, Lubrication Oil Temperature, and Coolant Temperature. These variables, measured in different units, help detect anomalies signaling potential malfunctions.

Mean values provide insight into typical conditions, but discrepancies between means and medians across all variables suggest extreme values or skewed distributions, indicating potential outliers or irregular engine behavior that may impact performance.

Table 1. Mean and Median

	Features	Mean	Median
0	Engine rpm	791.239263	746.000000
1	Lub oil pressure	3.303775	3.162035
2	Fuel pressure	6.655615	6.201720
3	Coolant pressure	2.335369	2.166883
4	lub oil temp	77.643420	76.817350
5	Coolant temp	78.427433	78.346662

The 95th percentile of each feature identify extreme values, with 5% of data points exceeding this threshold across all six variables. These high values suggest potential stress on engine components, indicating risks of overheating, excessive wear, or fuel inefficiencies, underscoring the need for proactive anomaly detection.

Table 2. percentage of variables data exceeding the 95th percentile.

	95th Percentile	Max Value	Range Beyond 95th Percentile	Percentage above 95th Percentile
Engine rpm	1324.00	2239.00	1324.00–2239.00	4.99%
Lub oil pressure	5.06	7.27	5.06–7.27	5.00%
Fuel pressure	12.21	21.14	12.21–21.14	5.00%
Coolant pressure	4.44	7.48	4.44–7.48	5.00%
lub oil temp	84.94	89.58	84.94–89.58	5.00%
Coolant temp	88.61	195.53	88.61–195.53	5.00%

Statistical Anomaly Detection

The Interquartile Range (IQR) method was used to define normal operating ranges for each engine variable. It calculates the 25th and 75th percentiles to set upper and lower bounds, flagging values outside these limits as anomalies. Engines with more than two anomalies across different variables are classified as potential malfunctions, enabling targeted inspection and maintenance.

Table 3. The percentage of outliers using IQR method.

```
Percentage of engines with 0 outliers: 76.27%
Percentage of engines with 1 outliers: 21.57%
Percentage of engines with 2 outliers: 2.10%
Percentage of engines with 3 outliers: 0.06%
3 was the highest number of outliers recorded per engine.
```

Using the Interquartile Range (IQR) method, 2.16% (422 engines) were identified to exhibiting anomalous behavior falling within the expected 1-5% range. This percentage suggests that the majority of engines are performing within normal parameters, while a small but meaningful subset requires further investigation. Flagging these outliers allows the company to proactively address potential faults before they escalate, optimizing maintenance efforts and minimizing operational risks.

Machine Learning Anomaly Detection

To detect anomalies in the ship's engine, two machine learning methods were used: Isolation Forest and One-Class SVM. These models, trained on historical sensor data, identify deviations from normal engine behavior to flag potential malfunctions.

One-Class SVM scales input data and learns a decision boundary around normal points, classifying anything outside as anomalous. It uses an RBF kernel to capture non-linear patterns, with the nu parameter set to 0.03, assuming up to 3% anomalies.

Isolation Forest isolates anomalies by randomly selecting features and splitting data. Since anomalies are less frequent, they require fewer splits. The model's contamination parameter was set to 0.03, and 100 estimators were used for efficiency.

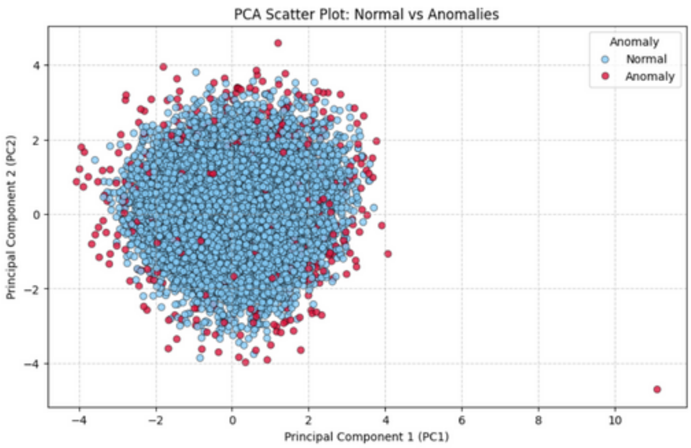
Principal Component Analysis (PCA) reduced data dimensionality, projecting it onto two principal components for better visualization. Anomalies detected by both models were plotted using PCA, making them easier to distinguish and aiding model evaluation.

Results

One-Class SVM Method

The One-Class SVM model, configured with a nu parameter of 0.03, identified 3.02% (589) of the engines as exhibiting anomalous behavior.

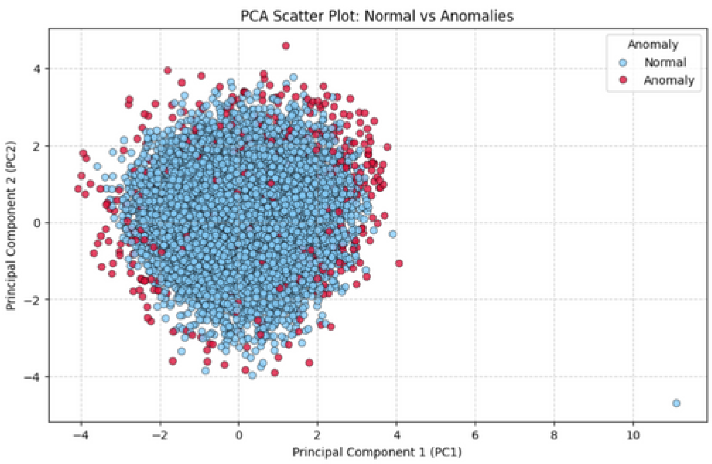
Figure 1. One-Class SVM method visualization using PCA plots



Isolation Forest Method

The Isolation Forest algorithm, configured with a contamination parameter of 0.03, identified 3.00% (587) of the engines as exhibiting anomalous behavior.

Figure 2. Isolation Forest method visualization using PCA plots



One-Class SVM Method vs Isolation Forest Method

*Both the One-Class SVM and Isolation Forest methods, when configured with matching 3% anomaly detection thresholds, identified approximately 3% of the dataset as anomalous. This consistency demonstrates that both models' outputs are strongly influenced by their respective hyperparameter settings. However, a notable distinction emerged in their sensitivity - the One-Class SVM flagged two additional anomalous points compared to Isolation Forest, suggesting it may offer slightly more sensitive detection capabilities for certain types of outliers.

The 2D scatter plot visualizations reveal important differences in how each model identifies anomalies. While both methods detected points in the periphery of the data distribution, they frequently selected different specific points as anomalous. The Isolation Forest showed a stronger tendency to identify points strictly along the outermost edges, whereas the One-Class SVM demonstrated a more nuanced approach, capturing some anomalies that appeared closer to the main data cluster. These patterns highlight the fundamental differences in their detection methodologies: Isolation Forest excels at identifying clear separation-based outliers, while One-Class SVM can detect more subtle, density-based anomalies. This complementary behavior suggests potential value in using both approaches for comprehensive anomaly detection in operational monitoring scenarios.

Conclusion and Recommendation

The analysis showed that method selection and threshold settings significantly impact anomaly classification. Since anomalies typically make up 1-5% of the data, all three methods— IQR (2.16%), One-Class SVM (3%), and Isolation Forest (3%)—proved effective in detecting engine faults. Their consistency highlights their reliability.

Recommendations:

- Deploy all three models to compare real-world performance, selecting the most accurate and interpretable one for long-term use. A hybrid approach could improve detection by cross-validating results.
- Regularly review and adjust anomaly thresholds to adapt to engine performance changes, ensuring timely fault detection while minimizing false positives.
- Consult domain experts to validate model results and refine detection based on engineering insights.
- Establish continuous monitoring and feedback to retrain models periodically, improving adaptability to new patterns.

Implementing these steps will enhance anomaly detection, minimize downtime, and ensure optimal engine performance, supporting operational efficiency and safety.

References

Devabrat, M., 2022. Predictive Maintenance on Ship's Main Engine using AI. Available at: https://dx.doi.org/10.21227/g3za-v415. [Accessed 5 March 2024]