

Overview and Verdict

In our analysis of predictive maintenance for a ship's main engine, we evaluated three methods to detect anomalies in engine performance: the Interquartile Range (IQR) method, Support Vector Machines (SVM), and Isolation Forest. Our findings indicate that the IQR method was not suitable for this task, as it either identified too few or too many outliers, making it unreliable. On the other hand, both SVM and Isolation Forest provided consistent results, identifying approximately 5% of the data points as outliers. Interestingly, SVM and IQR produced almost the same thresholds for detecting anomalies. Based on these results, we recommend that the company should monitor all features closely and pay particular attention when any of these features fall below the established thresholds. A crucial point of concern is when all features, one after another, pass their respective thresholds, signalling a critical alert.

Explanation

To identify anomalies in the engine performance data, we applied three different methods: IQR, SVM, and Isolation Forest. The dataset included various features such as engine rpm, lub oil pressure, fuel pressure, coolant pressure, lub oil temperature, and coolant temperature. These features are critical indicators of engine health, and detecting anomalies in these metrics is essential for predictive maintenance.

Interquartile Range (IQR) Method:

The IQR method is a common statistical technique used to identify outliers by determining the range within which the central portion of the data lies. It is calculated by finding the difference between the first quartile (Q1) and the third quartile (Q3) values of the data. Data points that fall below $Q1 - 1.5 * IQR$ or above $Q3 + 1.5 * IQR$ are considered outliers.

However, the IQR method proved to be ineffective for our dataset. When applied, it either detected an excessive number of outliers or too few, which compromised its reliability. This inconsistency suggests that the IQR method is not robust enough to handle the variability and complexity of engine performance data.

Support Vector Machines (SVM):

SVM is a powerful machine learning algorithm that can be used for classification and regression tasks. For anomaly detection, we used a one-class SVM, which is trained on the normal data and identifies data points that deviate significantly from this norm. The one-class SVM constructs a boundary around the normal data points, and any point outside this boundary is considered an anomaly (*fig. 1*).

The SVM method identified approximately 5% of the data as outliers, indicating that these points deviate from the normal operational patterns. This consistency in the detection rate suggests that SVM is a reliable method for identifying anomalies in engine performance data. The mean thresholds determined by SVM were as follows:

Engine rpm	791
Lub oil pressure	3,3
Fuel pressure	6,6
Coolant pressure	2,33
Lub oil temperature	78
Coolant temperature	78

Isolation Forest:

Isolation Forest is an ensemble learning method specifically designed for anomaly detection. It works by isolating observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. The process of isolating anomalies is more efficient because anomalies are fewer and different, thus requiring fewer splits to isolate.

Similar to SVM, the Isolation Forest method identified around 5% of the data as outliers. This alignment with the SVM results further supports the reliability of Isolation Forest in detecting anomalies within the engine performance data. The thresholds identified by Isolation Forest also aligned closely with those determined by SVM.

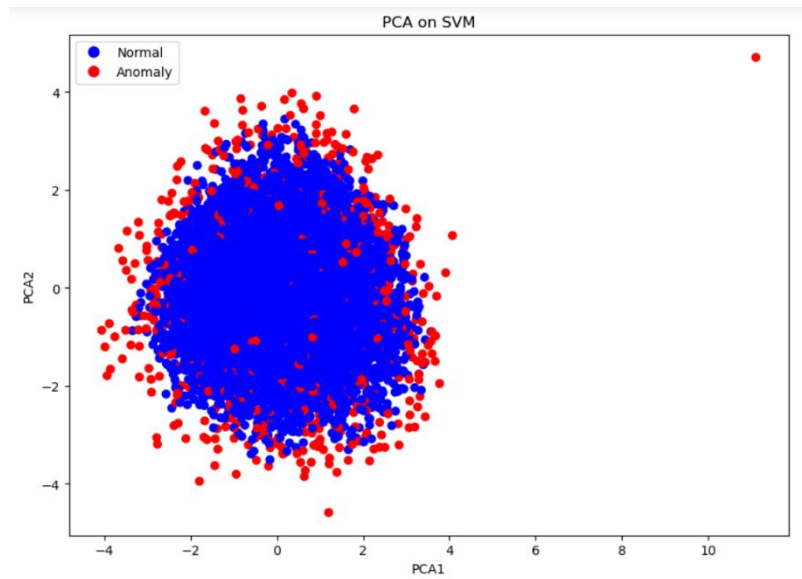
Conclusion

Based on our analysis, **we recommend** the following actions:

- 1) **Monitor All Features:** The company should continuously monitor all the features of engine performance. These features include engine rpm, lub oil pressure, fuel pressure, coolant pressure, lub oil temperature, and coolant temperature.
- 2) **Thresholds for Anomalies:** Special attention should be given to data points that fall below the established thresholds for each feature. The thresholds are determined based on the normal operational range of the engine parameters, and are as follows:
- 3) **Be particularly vigilant** when all features, one after another, pass their respective thresholds. This sequential crossing of thresholds is a critical alert that requires immediate attention to prevent potential failures.
- 4) **Regularly Update Data:** The SVM and Isolation Forest models should be regularly updated with new data to ensure they remain accurate and reflective of the current operational state of the engines.
- 5) **Investigate Detected Anomalies:** Any anomalies detected by SVM and Isolation Forest should be promptly investigated to prevent potential failures and ensure the longevity and reliability of the engine.

By implementing these recommendations, the company can enhance its predictive maintenance strategy, ensuring more reliable operation of its ship's main engine and reducing the risk of unexpected breakdowns.

APPENDIX



(fig.1)