Assignment 1 Mini-project: Detecting the anomalous activity of a ship's engine

Aadam Gafar, 14th June 2025

Introduction

Engine performance is critical to the shipping industry. In any fleet, poor performance can lead to inefficiencies, breakdowns, and delays, having a direct impact on profits, safety, and consumer satisfaction. This report details a system capable of detecting anomalous activity of a ship's engine via six key metrics: engine RPM, lubrication oil pressure, lubrication oil temperature, fuel pressure, coolant pressure, and coolant temperature. Identifying such cases will allow engineers to address issues before they arise, reducing the likelihood of the aforementioned failures and their effects on the bottom line.

Methodology

The engine's data consisted of six features: engine RPM, lubrication oil pressure, lubrication oil temperature, fuel pressure, coolant pressure, and coolant temperature. It had 19535 records for analysis. In total, three anomaly detection methods were compared: Interquartile Range (IQR), One-Class Support Vector Machine (SVM), and Isolation Forest.

As outlined in the project brief, it is reasonable to expect anomalies to comprise 1-5% of the total available data. In particular, the following advice was given: "*In isolation, a high RPM might not necessarily indicate a problem. [...] Anomalies should be flagged when certain combinations of features occur simultaneously. For example, if a high RPM coincides with high coolant pressure and elevated lubricant oil temperature, it might indicate a potentially problematic scenario that requires attention.*" [1] This guided the implementation of all three methods.

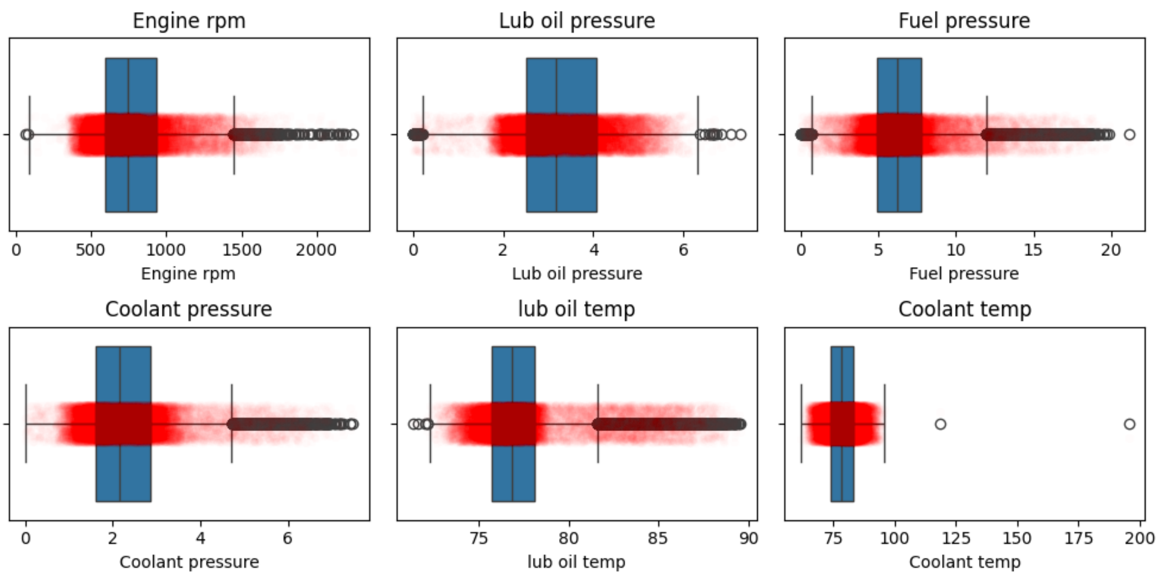
The foremost task was data pre-processing which involved ensuring the suitability, validity, and cleanliness of the data. It was inspected for null and duplicate values, of which it had neither. Descriptive statistics were computed and visualisations were created. Following this, each method was implemented.

- **IQR**: box plots, quartiles, and bounds were generated before tweaking parameters.
- **One-Class SVM**: the model was scaled and run with various parameter configurations. Dimensionality reduction was used before visualising results.
- **Isolation Forest**: the model was run with various parameter configurations. Dimensionality reduction was used before visualising results.

Analysis

After visualisations, the following features displayed a significant number of anomalies: engine RPM, fuel pressure, coolant pressure, and lubricant oil temperature. Specifically, these features had many outliers **above** their upper bound meaning that they are much **higher** than expected. The data described ship's engines that ran too fast, risked overheating, and endured excessive pressure.

Figure 1. - Comparing box plots of the dataset's six features



IQR

This method was run across multiple threshold values, to ascertain the most suitable level of stringency when classifying anomalies. The threshold defines how many features must be flagged as outliers before a record is considered anomalous. Results below:

- Threshold = 1, Anomalies = 23.73%
- Threshold = 2, Anomalies = 2.16%
- Threshold = 3, Anomalies = 0.06%

As can be seen, a threshold of 1 proves too loose in classifying anomalies, flagging nearly a quarter of the dataset as such. Conversely, threshold values of ≥3 are too stringent, flagging less than one percent of the dataset as such. A threshold value of 2, however, falls well within the expected and outlined range of 1-5% anomaly detection rate. **As such, we can deem any record with two or more anomalous features to be anomalous itself.**

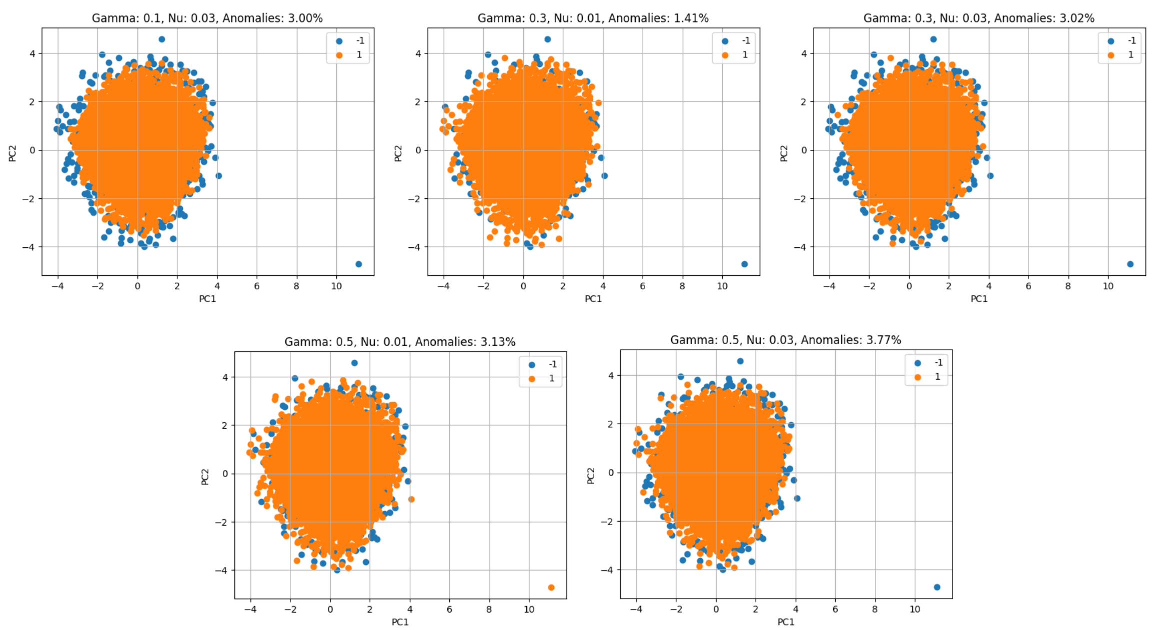
One-Class SVM

This model was also run across multiple parameter configurations, to ascertain the most suitable level of stringency when classifying anomalies. The parameters, gamma and nu, control, respectively: how strongly an individual data point affects the decision boundary, and the model's sensitivity to outliers. Results below:

- Gamma: 0.1, Nu: 0.03, Anomalies: 3.00%
- Gamma: 0.3, Nu: 0.01, Anomalies: 1.41%
- Gamma: 0.3, Nu: 0.03, Anomalies: 3.02%
- Gamma: 0.5, Nu: 0.01, Anomalies: 3.13%
- Gamma: 0.5, Nu: 0.03, Anomalies: 3.77%

As can be seen, there are multiple configurations suited to our goal of a 1-5% anomaly rate. To find the optimal configuration, we use PCA and visualise the results.

Figure 2. - Comparing One-Class SVM models at varying configurations

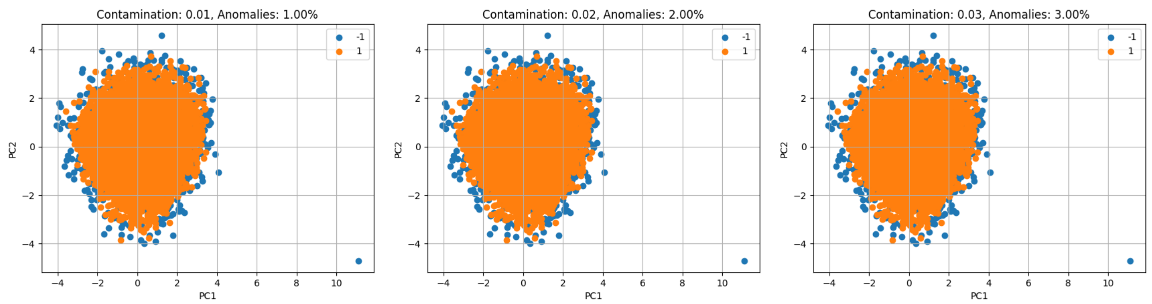


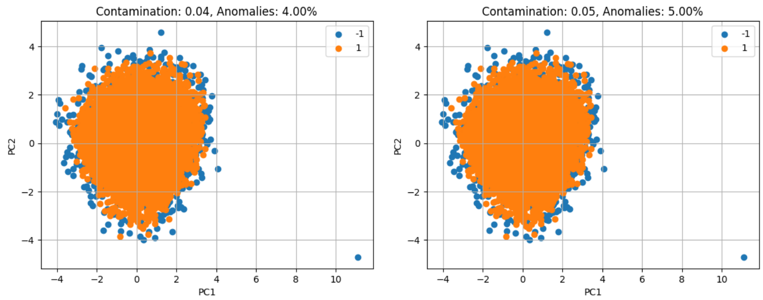
Our visuals should have clean separation between anomalous and regular data. This is best seen with the following configuration: Gamma: 0.1, Nu: 0.03. **Consequently, our model has an anomaly detection rate of 3.00%, well within our desired range of 1-5%.**

Isolation Forest

This model allows the anomaly detection rate to be directly influenced via a parameter known as the contamination factor. Our desired rate is 1-5% so this model was run across this range in increments of one percent, to ascertain the most suitable level of stringency when classifying anomalies. The visualisations are below:

Figure 3. Comparing Isolation Forest models at varying contamination rates



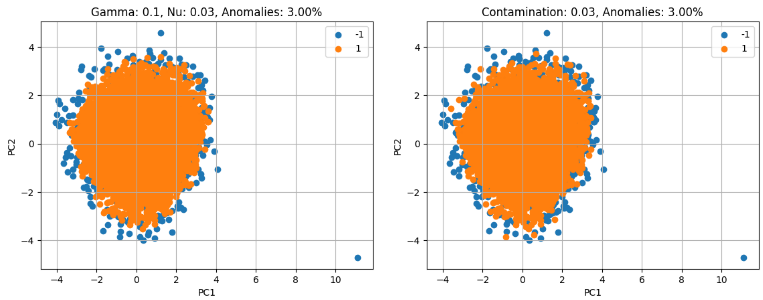


Visually, the plots are functionally identical, with little distinguishing features between them. **As a result, a contamination level of 0.03 can be considered optimal.** It strikes a balance between sensitivity and precision, lying squarely in the middle of the desired 1-5% anomaly detection rate.

Conclusion

After comparing methods, we have reached the desired 1-5% anomaly detection rate in all three methods. However, we may discard IQR on the basis that, unlike the algorithms, it is univariate and incapable of considering features in concert. The true decision lies in choosing One-Class SVM or Isolation Forest to detect anomalies.

Figure 4. - Comparing One-Class SVM (left) and Isolation Forest (right)



Reviewing the visualisations and the metrics, both models produce near identical results. **However, Isolation Forest can be considered superior for its simpler tweaking and robustness against larger datasets. It better aligns with the business need for accurate engine anomaly detection at scale.**

References

[1] FourthRev, 2025. *Assignment 1 Mini-project: Detecting the anomalous activity of a ship's engine*. Available at: https://fourthrev.instructure.com/courses/888/assignments/2934 (Accessed: 14 June 2025).