**5 Essential Qualities of Anomaly Detection Systems**

Ensuring your business is proactive and risk-proof.



Photo by [Will Myers](https://unsplash.com/@will_myers?utm_source=medium&utm_medium=referral" \t "_blank) on [Unsplash](https://unsplash.com/?utm_source=medium&utm_medium=referral" \t "_blank)

**Introduction**

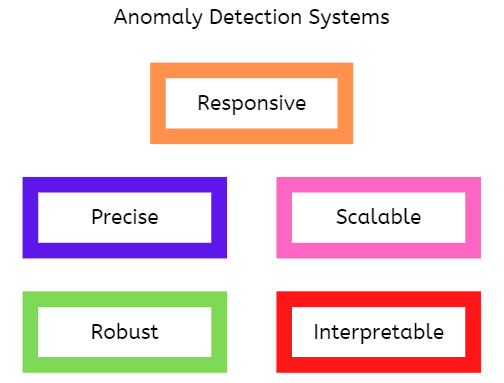
This article covers the five essential qualities of anomaly detection systems that you should know as a data practitioner.

Anomaly detection is a critical application of data science that offers excellent value in protecting businesses and creating proactiveness against threats. It can also help to detect the onset of faults in complex machines such as wind turbines through condition monitoring.

Anomaly detection solutions can range from setting simple statistical thresholds based on domain knowledge to building sophisticated machine-learning systems for real-time cases. For example, when the data is normally distributed, data points falling outside 2 to 3 standard deviations from the mean may be extracted as anomalies for further investigation.

Of course, there are tradeoffs in the different approaches to building anomaly detection systems and not all of them offer the same solution quality.

**5 Essential Qualities of Anomaly Detection Systems**



Essential qualities of anomaly detection systems (image by author)

Now, let’s discuss the five main attributes of an excellent anomaly detection system.

1. **Responsiveness**

When anomaly detection solutions are built for threat detection in critical areas such as power and transport systems, they must be quick to act to prevent severe damage.

In many real-world applications, anomaly detection results are most useful before the damage is done.

Many implementations and tutorials on anomaly detection models are based on static or historical data which does not consider the complexity of responsiveness. However, it can make or mar the success of your project in real-world applications.

For example, a quicker response from anomaly detection systems will enable businesses to have a minimal response time to platform glitches resulting in an overall improvement in customer satisfaction.

*Helpful tip:*Parallelizing modeling tasks can help to speed up the responsiveness of a system in real-time applications, especially in cases where online machine learning is required. In addition, offline learning can also help reduce computing time during inference.

**2. Precision**

Too many false alarms within a short period will only make stakeholders ignore your system’s alerts and this renders the solution useless for any practical application.

Fool me once, it’s on you but if you fool me twice it’s on me.

Hence, anomaly detection systems need to have great precision while having satisfactory [sensitivity](https://deepchecks.com/glossary/sensitivity-and-specificity-of-machine-learning/" \l ":~:text=Sensitivity%20in%20Machine%20Learning%20can,positives%20of%20each%20available%20category." \t "_blank).

Anomalies are typically rare and models may overcompensate for this during training. For instance, false alarms may result when the model is too sensitive to anomalies. Hence, classifying normal operational data as anomalies in the process.

*Helpful tips:*In classification tasks, this issue may be addressed during training by selecting the proper metrics that address precision issues such as F1 score and area under the receiver operating characteristic (ROC) curve. For regression tasks, precision may be increased by raising the detection threshold thereby alerting stakeholders of only the most severe anomalies.

**3. Scalability**

The goal of every business is to grow. Hence, integrated systems such as those for anomaly detection must be scalable to meet business goals. In addition, the responsiveness of a system may be impacted at scale if this is not taken into consideration during the design stage. Computational cost may also rise significantly at large-scale deployment making the solution impractical.

For instance, a simpler and less accurate solution may be preferred to a more complex and marginally more accurate solution that is less scalable to handle more data.

*Helpful tips:*Computing overhead may be addressed with code reformatting and parallelization where possible. In addition, complex techniques may be replaced with more computationally efficient but equally accurate methods.

**4. Robustness**

A robust anomaly detection system must be able to handle data quality issues such as missing values, especially in real-time applications. The system should also be able to handle bad examples that can bias the model during training.

For example, some features may not be available at inference time. A robust anomaly detection system needs to be able to handle such cases.

*Helpful tips:*Issues with robustness can be addressed in the data pipeline to properly capture the different data quality concerns. In addition, some algorithms (e.g random forests) that are robust to missing values and outliers can be used.

**5. Interpretability**

For stakeholders to act proactively, anomaly detection results need to be interpretable. This is key to building more resilient business processes. In addition, interpretable results help engineers during [root cause analysis](https://asq.org/quality-resources/root-cause-analysis" \t "_blank).

A common machine learning dilemma may arise where the more sophisticated and accurate models are less interpretable while simpler and less accurate models give better explainable results.

For instance, interpretability may be lost during feature engineering when using techniques such as principal component analysis (PCA) to reduce data dimensionality.

*Helpful tips:* Complex and simple ML solutions may be developed where the former is used for accurately detecting the anomalies and the latter is used for providing explanations as necessary. Replacing methods that reduce the interpretability of results may also be helpful.

**Conclusions**

In this article, we covered the essential qualities of anomaly detection systems: responsiveness, precision, scalability, robustness, and interpretability. These attributes must be considered during the design stage to derive the best value from your data science project.