

The Beginners Guide to Anomaly Detection

Table of Contents

- 01 Why Is Anomaly Detection Important?
- 02 What Are Anomalies?
- 03 Common Anomaly Detection Use Cases
- 04 Fighting Financial Crime
- 05 Monitoring Equipment Sensors
- 06 Healthcare Claims Fraud
- 07 Manufacturing Defects
- 08 More Use Cases
- 09 Techniques for Anomaly Detection
- 10 Conclusion



01 Why Is Anomaly Detection Important?

Today, massive quantities of business operations data are generated daily. That data can help gain a competitive advantage, but only if used correctly. One way is through anomaly detection, which can stop a minor issue from becoming a widespread, time-consuming problem. In this guide, we will cover anomaly detection basics, common use cases, and some key techniques to keep in mind.

02 What Are Anomalies?

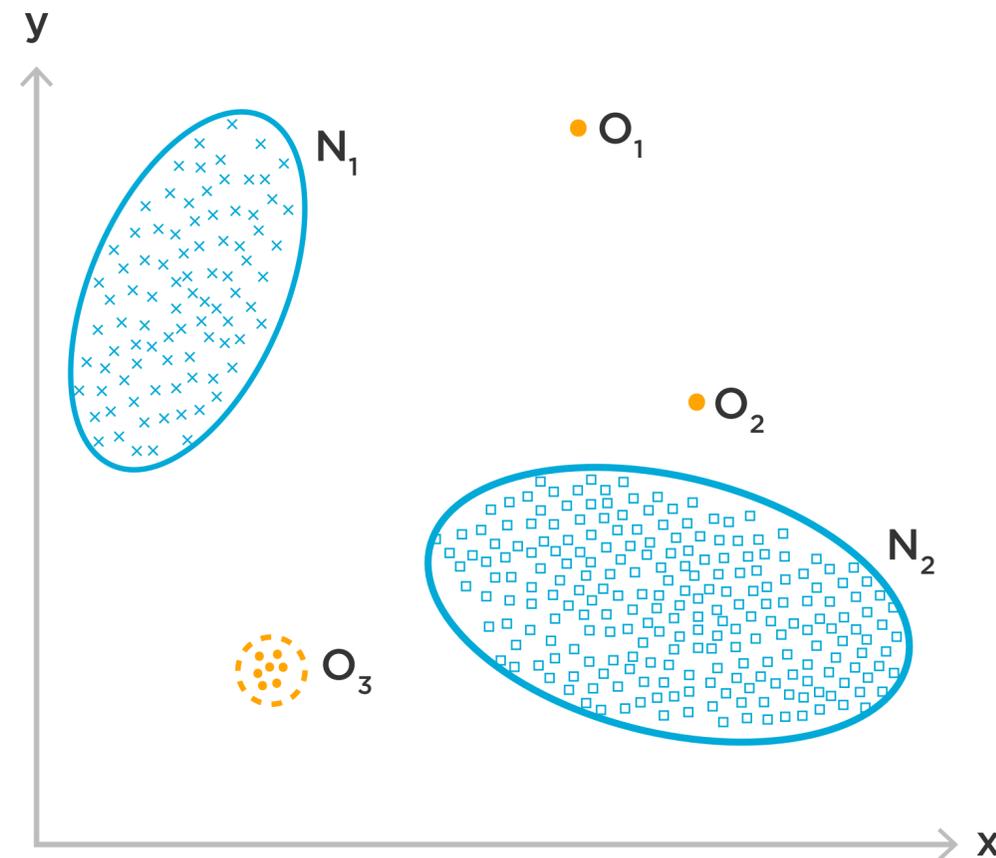
Before we dive in, let's take a step back. What exactly are anomalies?

An **anomaly** is an unexpected change or deviation from the expected pattern in a dataset. Therefore, anomaly detection is a way of detecting abnormal behavior. It's important to note that anomalies aren't necessarily good or bad, but companies should be alerted to any break in pattern to assess whether actions need to be taken.

What Is the Difference between Anomalies and Outliers?

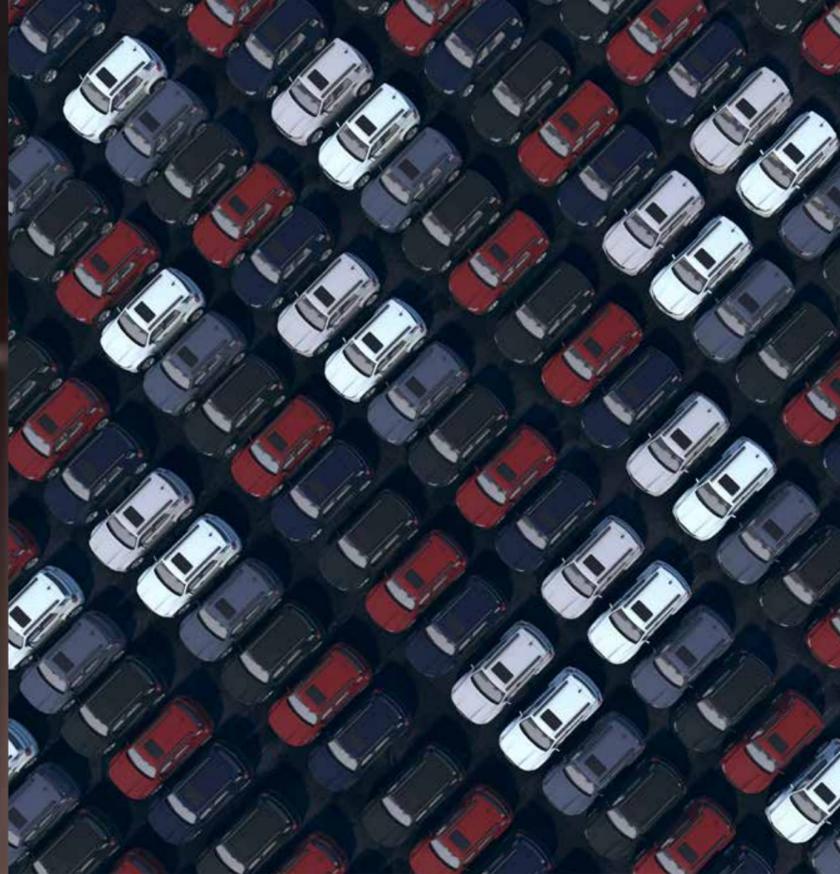
There is much debate on this topic, and many people use the terms interchangeably. Synonyms for outliers may include “discordant observations, exceptions, aberrations, surprises, peculiarities, or contaminants in different application domains.”¹ We take the position that anomalies are similar, but not identical, to outliers.

Assuming that all data is generated by a set of processes, outliers are points with a low probability of occurrence within a given dataset generated by those processes. They are observation points that are distant from other observations within the normal population; However, they don't necessarily represent abnormal behavior or behavior that occurred because of a different process. Outliers are generated by the same process but occur with lower probability, whereas anomalies are patterns that are generated by different processes.



A simple example of anomalies (o_1 , o_2 , o_3) in a 2D dataset

¹ Banerjee, A. (2009, July). Anomaly Detection: A Survey. Retrieved from <http://cinslab.com/wp-content/uploads/2019/03/xiaorong.pdf>.



03 Common Anomaly Detection Use Cases

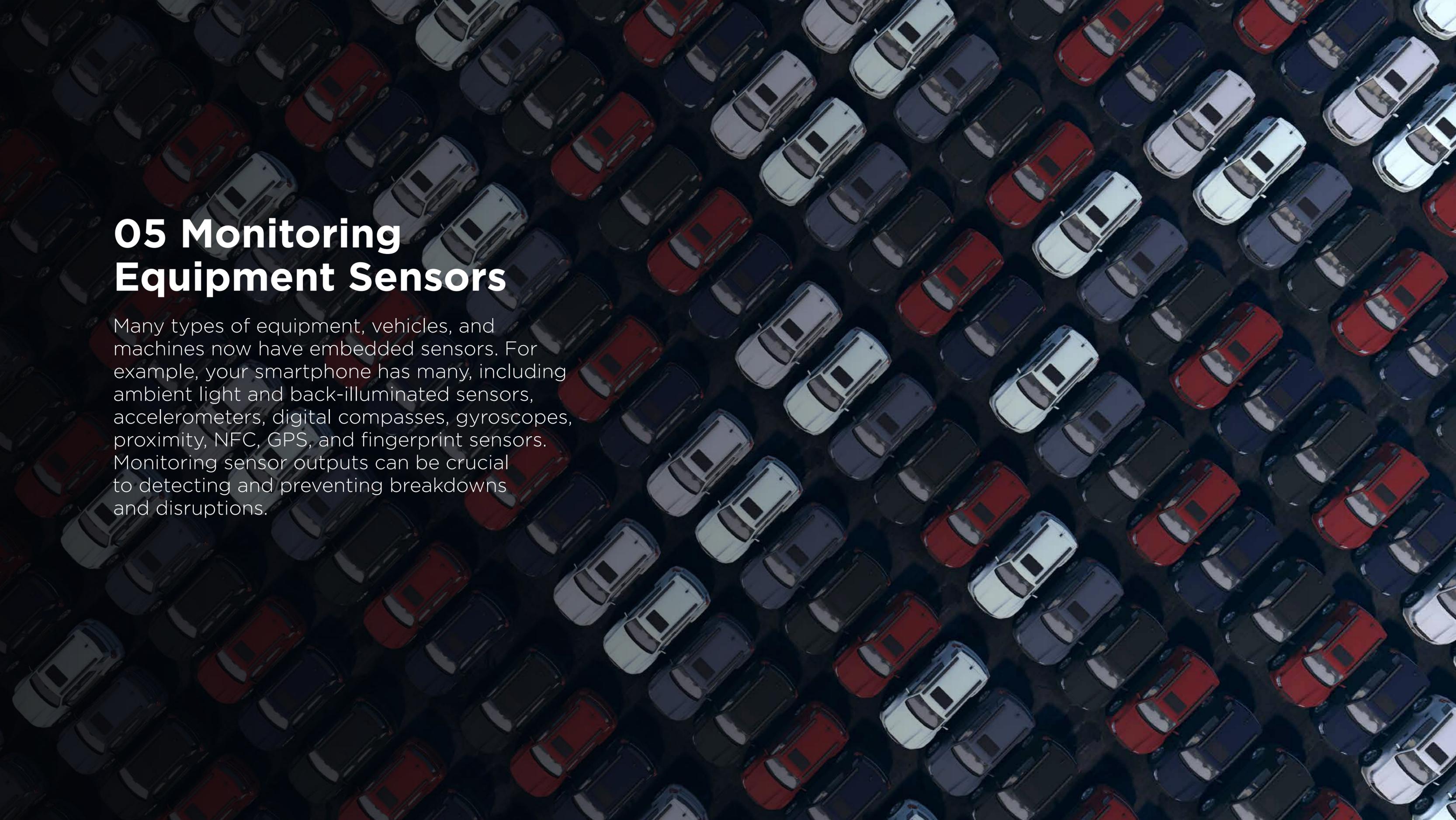
Across nearly every industry, there are many important business use cases for anomaly detection. A few of the most common:

- Financial crimes
- Equipment sensors
- Healthcare fraud
- Manufacturing defects

04 Fighting Financial Crime

In finance, trillions of dollars worth of transactions execute every minute. Identifying suspicious transactions in real time can provide organizations with a competitive edge. To identify abnormal transactions, clients, and suppliers, leading financial companies have increasingly adopted big data analytics, including machine learning techniques, to detect anomalies among the voluminous sea of data being generated.



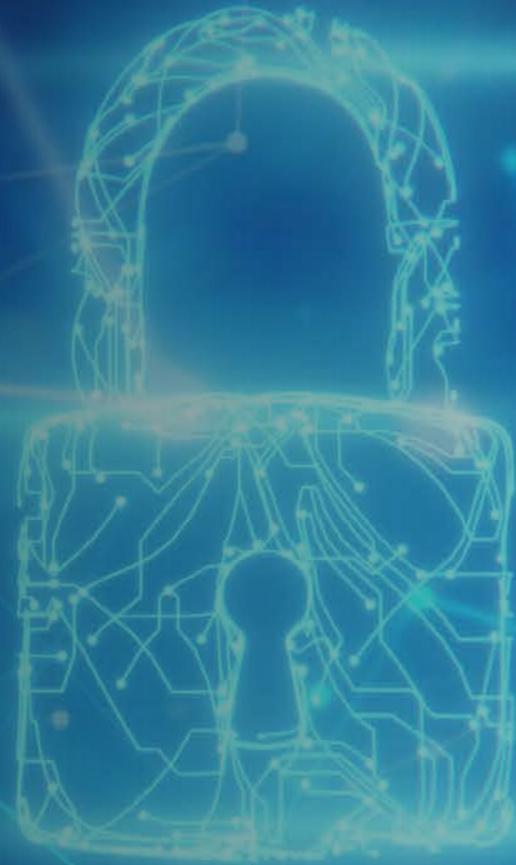


05 Monitoring Equipment Sensors

Many types of equipment, vehicles, and machines now have embedded sensors. For example, your smartphone has many, including ambient light and back-illuminated sensors, accelerometers, digital compasses, gyroscopes, proximity, NFC, GPS, and fingerprint sensors. Monitoring sensor outputs can be crucial to detecting and preventing breakdowns and disruptions.

06 Healthcare Claims Fraud

Insurance fraud is common in the healthcare industry. By some estimates, U.S. healthcare fraud costs the nation about \$68 billion annually.¹ It is vital for insurance companies to identify fraudulent claims and ensure that no payout is made. In the past few years, many companies have invested heavily in big data analytics to build supervised, unsupervised, and semi-supervised models to detect insurance fraud.



¹ The National Health Care Anti-Fraud Association. (2017, March). Fraud Statistics. Retrieved from <https://www.bcbsm.com/health-care-fraud/fraud-statistics.html>.

07 Manufacturing Defects

Some companies continuously monitor sensor data on manufactured components with an autoencoder model. As the model scores new data, any defects (anomalies) are quickly detected and preventative actions can be taken.





08 More Use Cases

Beyond the most common use cases already described, anomaly detection can be applied across a wide variety of industries:

- Military surveillance: Image recognition
- Cybersecurity: Intrusion detection
- Safety systems: Fault detection
- Hacking protection: Anomalous network traffic detection
- Weather: Heat wave or cold snap implications
- MRI imaging: Alzheimer's or malignant tumor implication
- Spacecraft sensors: Faulty component identification



09 Techniques for Anomaly Detection

1. Visual Discovery

Data or business analysts build data visualizations to find unexpected behavior, often requiring prior business knowledge and creative thinking to find the answers with the right data visualizations.

2. Supervised Learning

Persons with business knowledge in a particular industry label a set of data points as normal or anomalous. An analyst then uses this labeled data to build machine learning models to predict anomalies on unlabeled new data.

3. Unsupervised Learning

Unlabeled data is used to build unsupervised machine learning models that are then used to predict new data. Since the model is tailored to fit normal data, the small number of data points that are anomalous stand out.

10 Conclusion

With an overwhelming amount of data being collected today, businesses must be able to track patterns and, more importantly, detect anomalies. Detecting anomalies in data patterns can help businesses uncover actionable insights and become more efficient and competitive in the digital age.

With TIBCO Data Science software, organizations can use machine learning models to specify expected behavior, monitor new data, and find unexpected behavior.

Now that you have a handle on the basics, learn more about how to get ahead with anomaly detection at www.tibco.com/solutions/anomaly-detection.