**ANOMALY DETECTION :**

**Turi Use cases :**

<https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings>

<https://data.seattle.gov/Transportation/Fremont-Bridge-Hourly-Bicycle-Counts-by-Month-Octo/65db-xm6k>

<https://fred.stlouisfed.org/series/DJIA>

Anomaly detection is a concept widely applied to numerous domains. Several techniques of anomaly detection have been developed over the years, in practice as well as research. The application of this concept has extended to diverse areas, from ***network intrusion detection*** to ***novelty detection in robot behavior***. In the business world, the application of these techniques to ***fraud detection*** is of a special interest, driven by the great losses companies endure because of such fraudulent activities. The use of ML in fraud detection or law enforcement is all about identifying unusual events as potential threats.

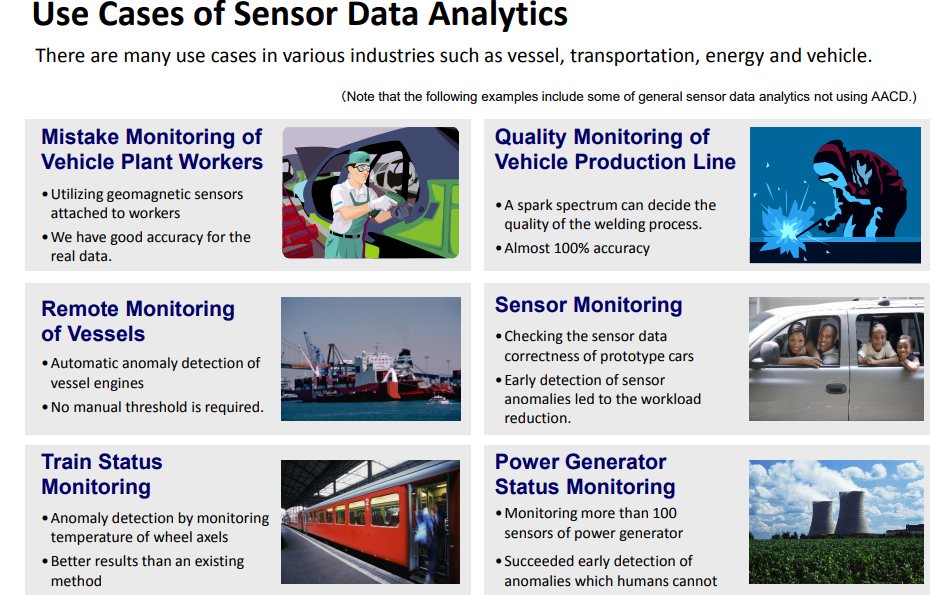
Anomaly detection is similar to — but not entirely the same as — noise removal and novelty detection. **Novelty detection** is concerned with identifying an unobserved pattern in new observations not included in training data — like a sudden interest in a new channel on YouTube during Christmas, for instance. **Noise removal** ([NR](http://datamining.rutgers.edu/publication/tkdehcleaner.pdf)) is the process of immunizing analysis from the occurrence of unwanted observations; in other words, removing noise from an otherwise meaningful signal.

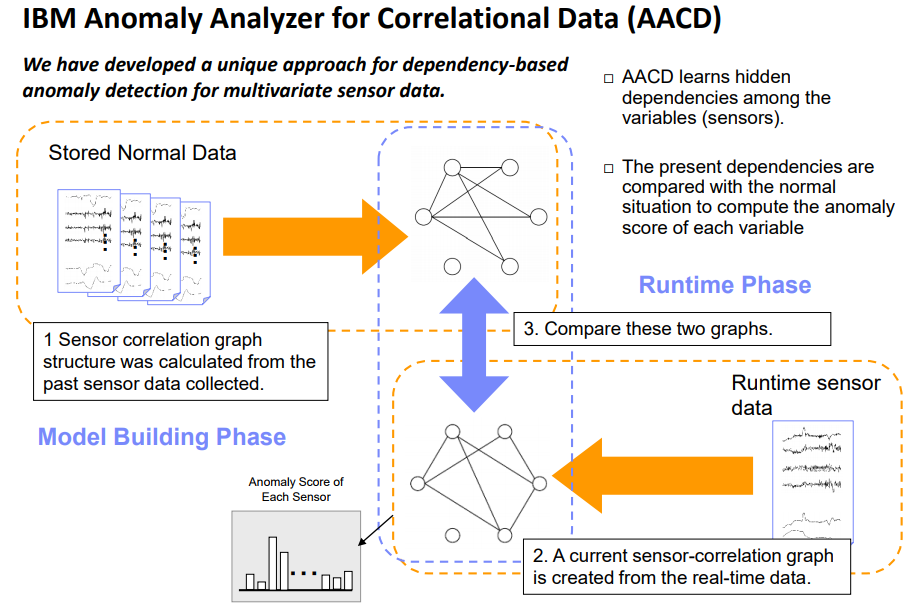
**APPLICATIONS :**

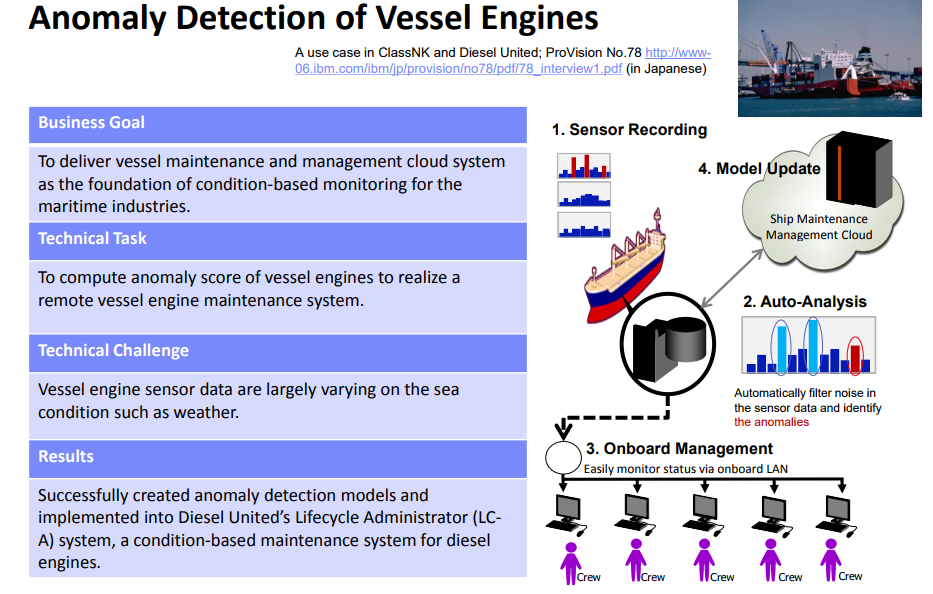
* Dealing with **data anomalies**: 1) when choosing a suitable learning algorithm or cleansing the input accordingly (include the anomaly, vs include it but label it as an outlier, vs exclude it as noise), or 2) when they confound the interpretation that you've proposed (that failed to explain / predict those anomalies).
* Detection of anomalies (outliers or rare events) has recently gained a lot of attention in many **security domains**, ranging from video surveillance and security systems to intrusion detection and fraudulent transactions. For example, in ***video surveillance applications, video trajectories*** that represent suspicious and/or unlawful activities (e.g. identification of traffic violators on the road, detection of suspicious activities in the vicinity of objects) represent only a small portion of all video trajectories. Similarly, in the **network intrusion detection domain**, the number of cyber attacks on the network is typically a very small fraction of the total network traffic,*for instance*, an employee who is accidentally or intentionally leaking large amounts of data outside the company intranet. Or maybe a hacker opening connections on non-common ports and/or protocols. In the specific case of Internet security, anomaly detection could be used for stopping new malware from spreading out by simply looking at spikes of visitors on non-trusted domains. And even if cyber security is not your core business, you should protect your network with data-driven solutions.
* Manual threshold setting is very difficult to consider all dependencies among sensors

□ Accuracy of threshold-based method is poor – Many false alerts and miss alerts

□ In various dynamic systems, threshold is changing dynamically – i.e. high temp. under high pressure is not anomaly, but high temp. under low pressure is anomaly.







**USE CASE :**

<https://www.datameer.com/usecases/datasources/useCase/Network-Security-Anomaly-Detection/>

<https://www.datascience.com/blog/python-anomaly-detection>