Institución Universitaria
Reacreditada en Alta Calidad

Anomaly Detection

Ingeniería Electrónica

Somos Innovación Tecnológica con Sentido Humano

Alcaldía de Medellín

Definition

Anomaly detection is any process that finds the outliers of a dataset; those items that don't belong. These anomalies might point to unusual network traffic, uncover a sensor on the fritz, or simply identify data for cleaning, before analysis.

In today's world of distributed systems, managing and monitoring the system's performance is crucial. With hundreds or thousands of items to watch, anomaly detection can help point out where an error is occurring, enhancing root cause analysis and quickly getting tech support on the issue. Anomaly detection helps the monitoring cause of chaos engineering by detecting outliers, and informing the responsible parties to act.

https://www.bmc.com/blogs/machine-learning-anomaly-detection/

Applications

In enterprise IT, anomaly detection is commonly used for:

- Data cleaning
- Intrusion detection
- Fraud detection
- Systems health monitoring
- Event detection in sensor networks
- Ecosystem disturbances

https://www.bmc.com/blogs/machine-learning-anomaly-detection/

Anomaly Detection with ML

> *"Anomaly detection (AD) systems are either manually built by experts setting thresholds on data or constructed automatically by learning from the available data through machine learning (ML)." Bram Steenwinckel*

Machine learning, then, suits the engineer's purpose to create an AD system that:

- Works better
- Is adaptive and on time
- Handles large datasets

Despite these benefits, anomaly detection with machine learning can only work under certain conditions.

https://www.bmc.com/blogs/machine-learning-anomaly-detection/

Anonaly Detection Settings

In a 2018 lecture, Dr. Thomas Dietterich and his team at Oregon State University explain how anomaly detection will occur under three different settings. They all depend on the condition of the data. The three settings are:

- Supervised
- Clean
- Unsupervised

https://www.bmc.com/blogs/machine-learning-anomaly-detection/

Anonaly Detection - Supervised

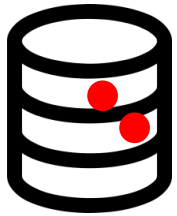
Nomimal Anomaly

Popular ML algorithms for structured data:

- Support vector machine learning
- k-nearest neighbors (KNN)
- Bayesian networks
- Decision trees

https://www.bmc.com/blogs/machine-learning-anomaly-detection/

Anonaly Detection - Clean



Anomaly

Nomimal

In the Clean setting, all data are assumed to be “nominal”, and it is contaminated with “anomaly” points.
The clean setting is a less-ideal case where a bunch of data is presented to the modeler, and it is clean and complete, but all data are presumed to be nominal data points. Then, it is up to the modeler to detect the anomalies inside of this dataset.

https://www.bmc.com/blogs/machine-learning-anomaly-detection/

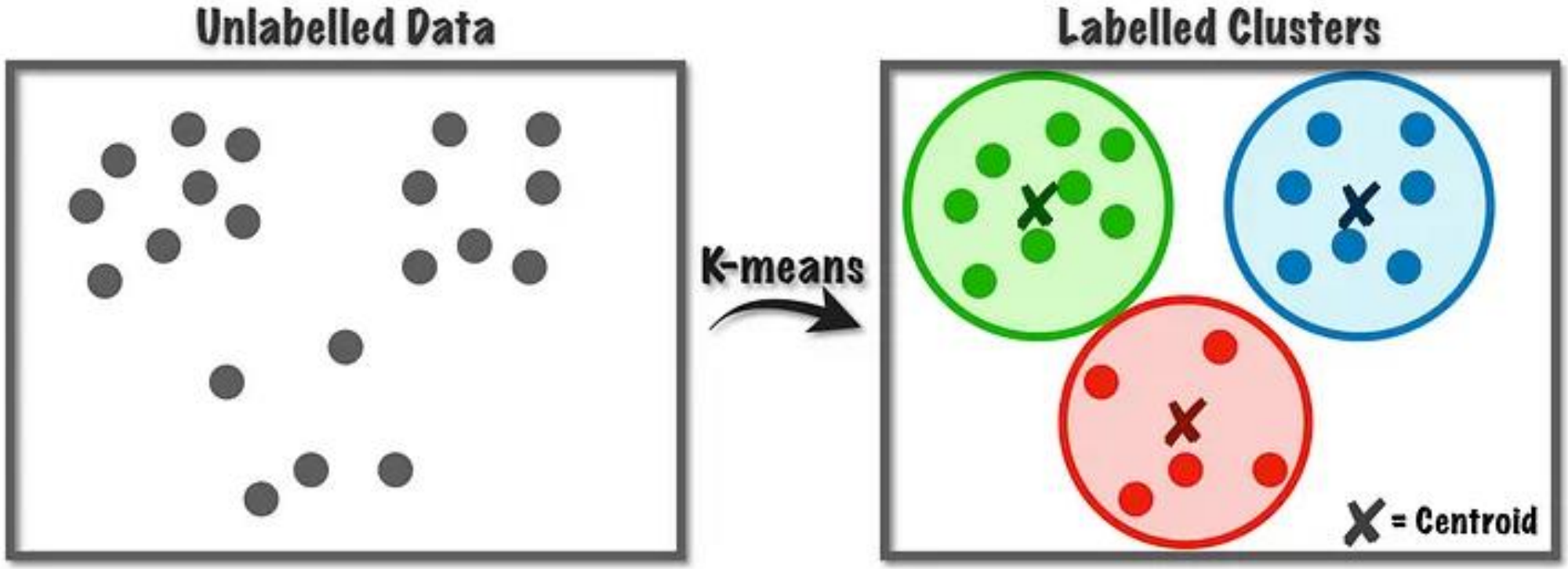
Anonaly Detection - Clean

Nomimal and Anomaly datapoints

The datasets in the unsupervised case do not have their parts labeled as nominal or anomalous. There is no ground truth from which to expect the outcome to be. The model must show the modeler what is anomalous and what is nominal.

https://www.bmc.com/blogs/machine-learning-anomaly-detection/

K-means



https://towardsdatascience.com/k-means-a-complete-introduction-1702af9cd8c

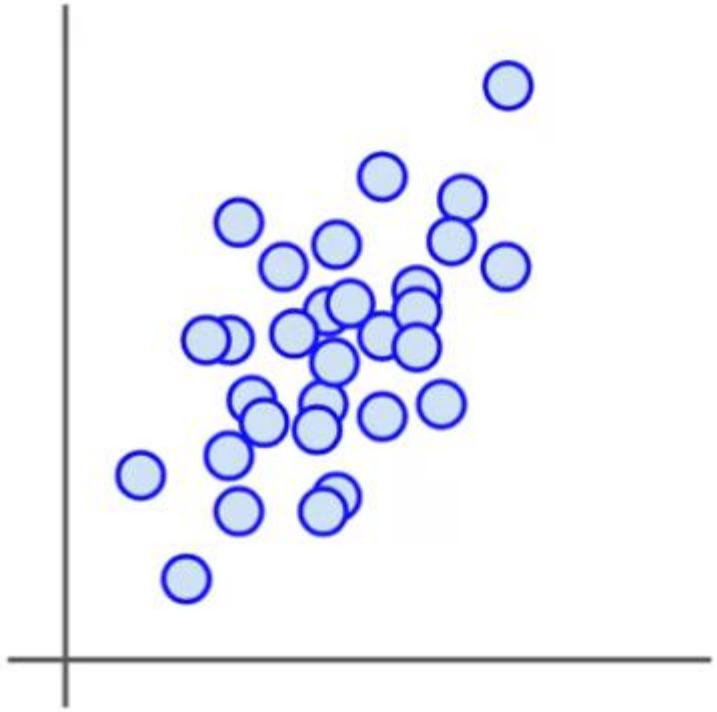
K-means

Original image

$K = 2$

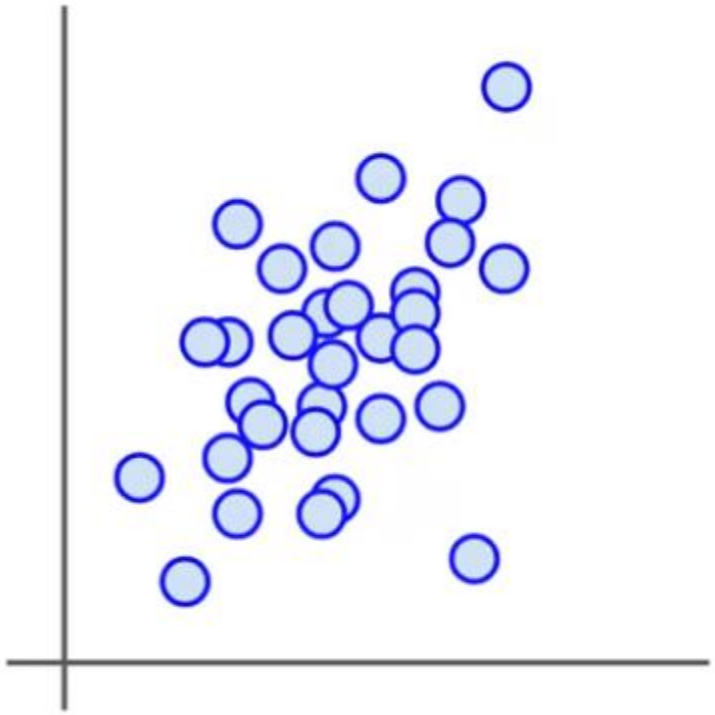
$K = 3$

Conceptual example

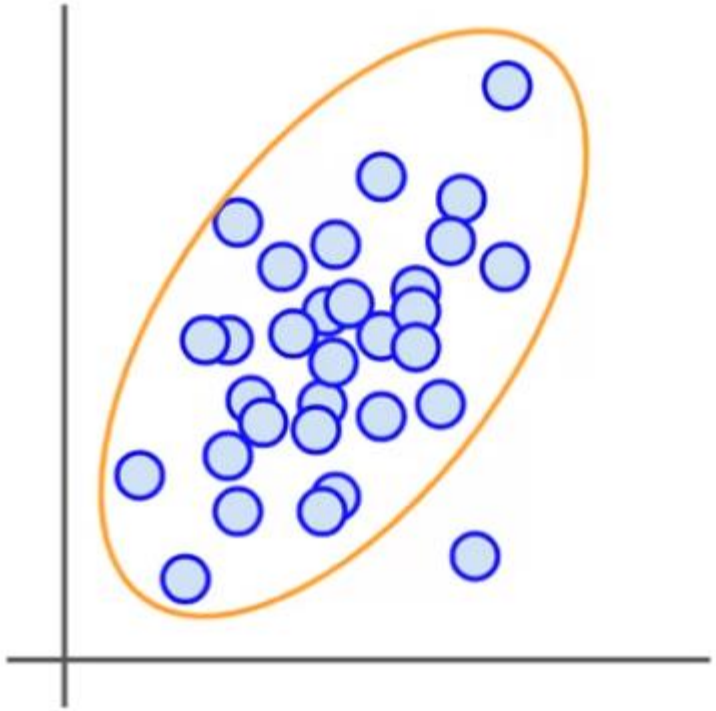


https://www.coursera.org/learn/introduction-to-embedded-machine-learning/lecture/nqjHZ/anomaly-detection

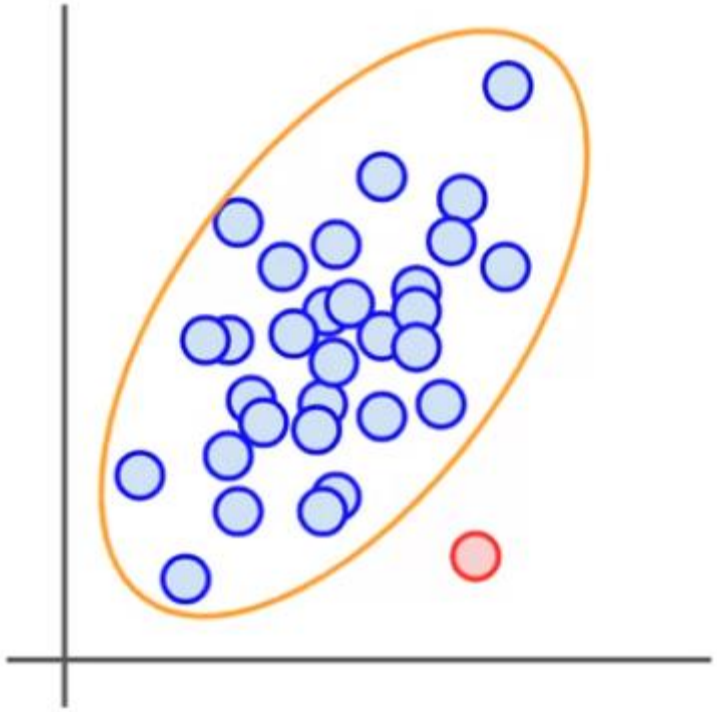
Conceptual example



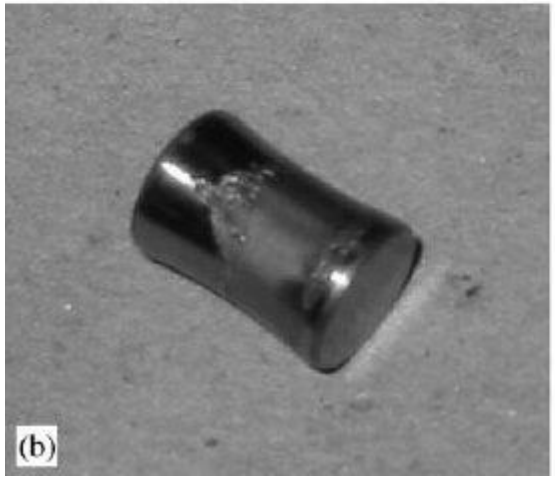
Conceptual example



Conceptual example

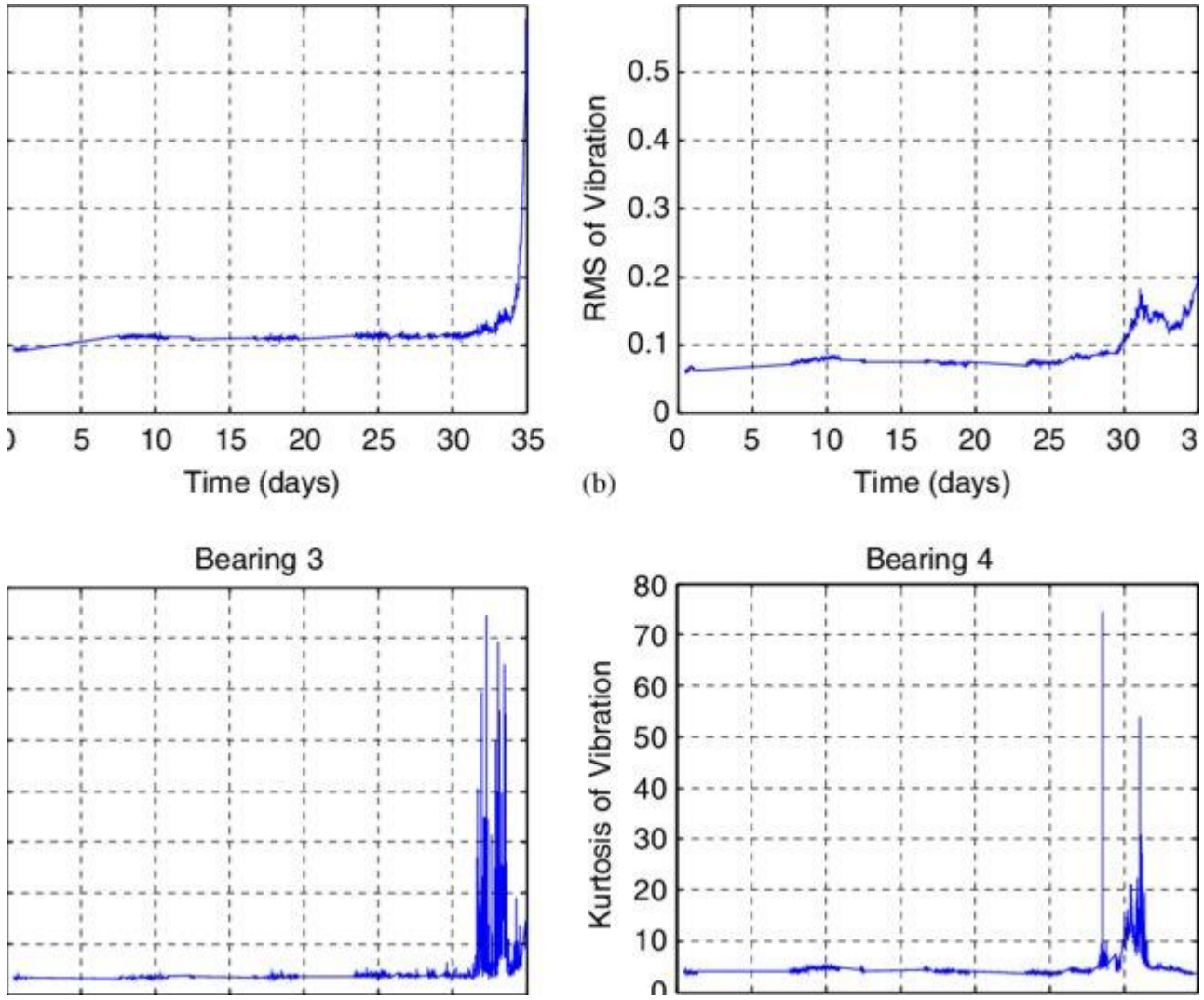


Bearing Anomaly



https://www.researchgate.net/publication/223556476_Wavelet_filter-based_weak_signature_detection_method_and_its_application_on_rolling_element_bearing_prognostics/figures}

Bearing Anomaly



Institución Universitaria
Reacreditada en Alta Calidad

¡Gracias!

Somos Innovación Tecnológica con Sentido Humano

Alcaldía de Medellín