ARUNDO

Applied Machine Learning for Anomaly Detection on Equipment

Alexandra Gunderson
Lukasz Mentel
Trung Doan

Alexandra

Lukasz

Trung

Copyright © 2018 Arundo Analytics, Inc. All rights reserved
ARUNDO

9:00 - 9:30	Welcome & Introduction to Anomaly Detection
9:30 - 10:00	Set-up environment
10:00 - 10:30	Walk through examples of AD methods in Jupyter notebooks
10:30 - 12:00	Improve models

-------- LUNCH BREAK --------

13:30 - 14:00	Introduction to deployment
14:00 - 15:30	Continue to improve model & deploy to cloud
15:30 - 16:00	Make sure final model is deployed
16:00 - 16:30	Review the results and wrap-up

Copyright © 2018 Arundo Analytics, Inc. All rights reserved
ARUNDO

BUZZWORD BINGO

Digital Transformation	Operational Intelligence	Streaming Analytics
IT Operations Analytics	**Industry 4.0**	Blended Analytics

Copyright © 2018 Arundo Analytics, Inc. All rights reserved

ARUNDO

Increase efficiency and productivity

Data is the jetfuel

Copyright © 2018 Arundo Analytics, Inc. All rights reserved
ARUNDO

Decrease downtime

Unexpected downtime on a single asset can cost upwards of a million dollars per day

INTERCONNECTED AMBITIONS

Increase efficiency

Increase profits despite a decreasing price per barrel

Scalable, actionable insight

Make models which can be easily applied to the company’s entire portfolio

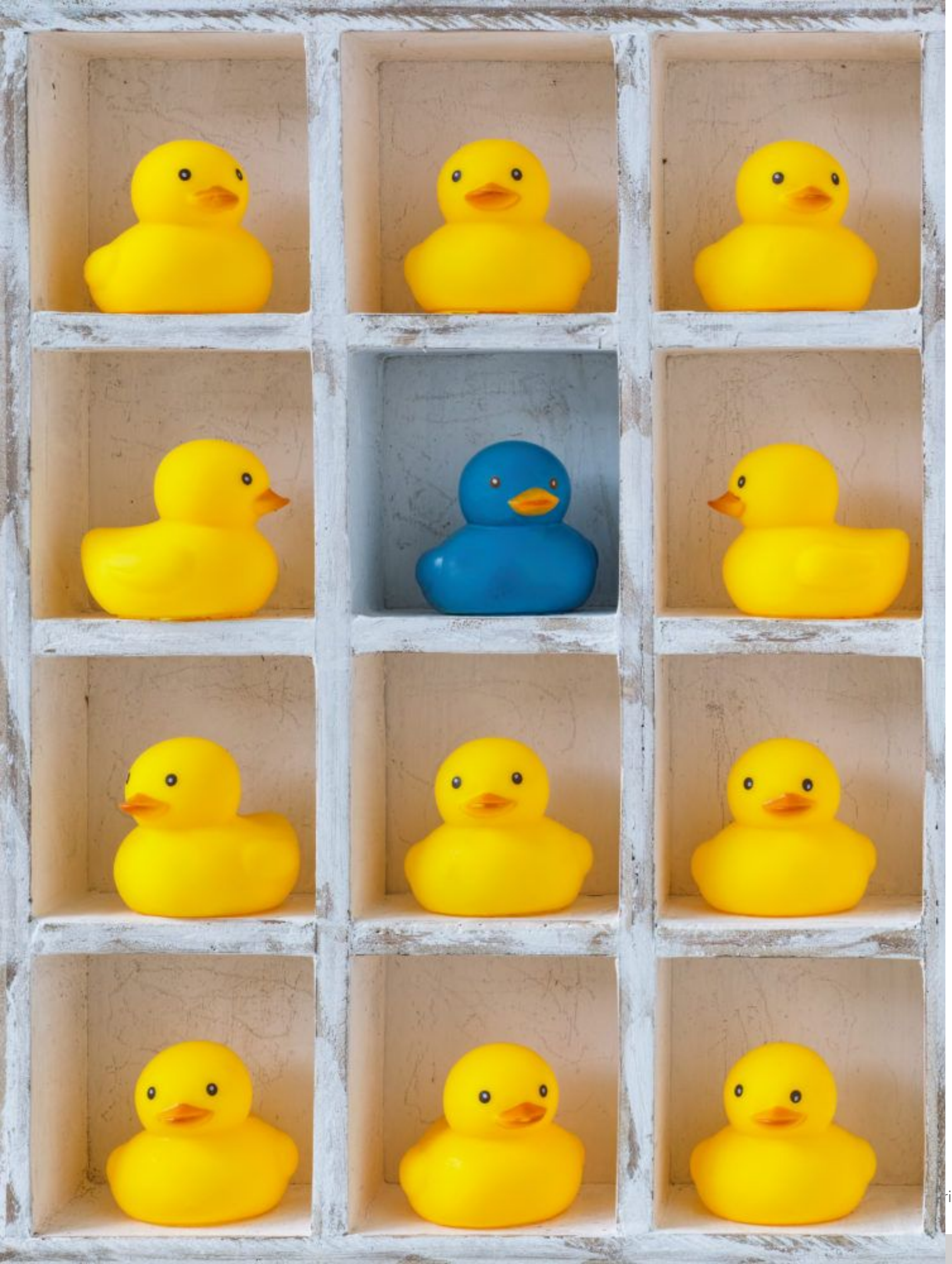
Copyright © 2017 Arundo Analytics, Inc. All rights reserved

ARUNDO

provides software products to **enable** enterprise-scale machine learning and advanced analytics applications for **industrial companies**

Copyright © 2018 Arundo Analytics, Inc. All rights reserved


What is an anomaly?

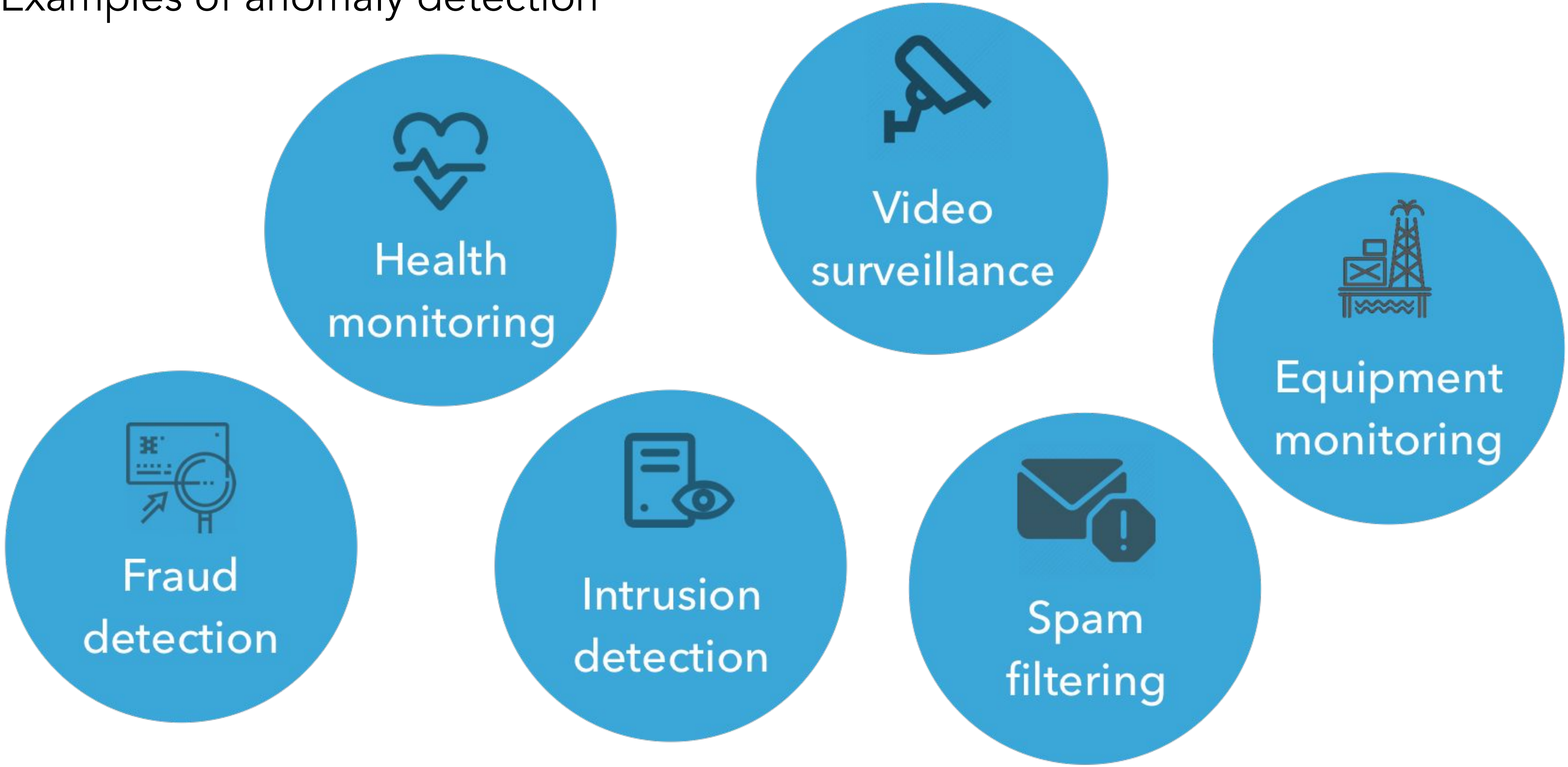


"DATAPOINTS, ITEMS, OBSERVATIONS OR EVENTS THAT DO NOT CONFORM TO THE EXPECTED PATTERN"

right © 2017 Arundo Analytics, Inc. All rights reserved

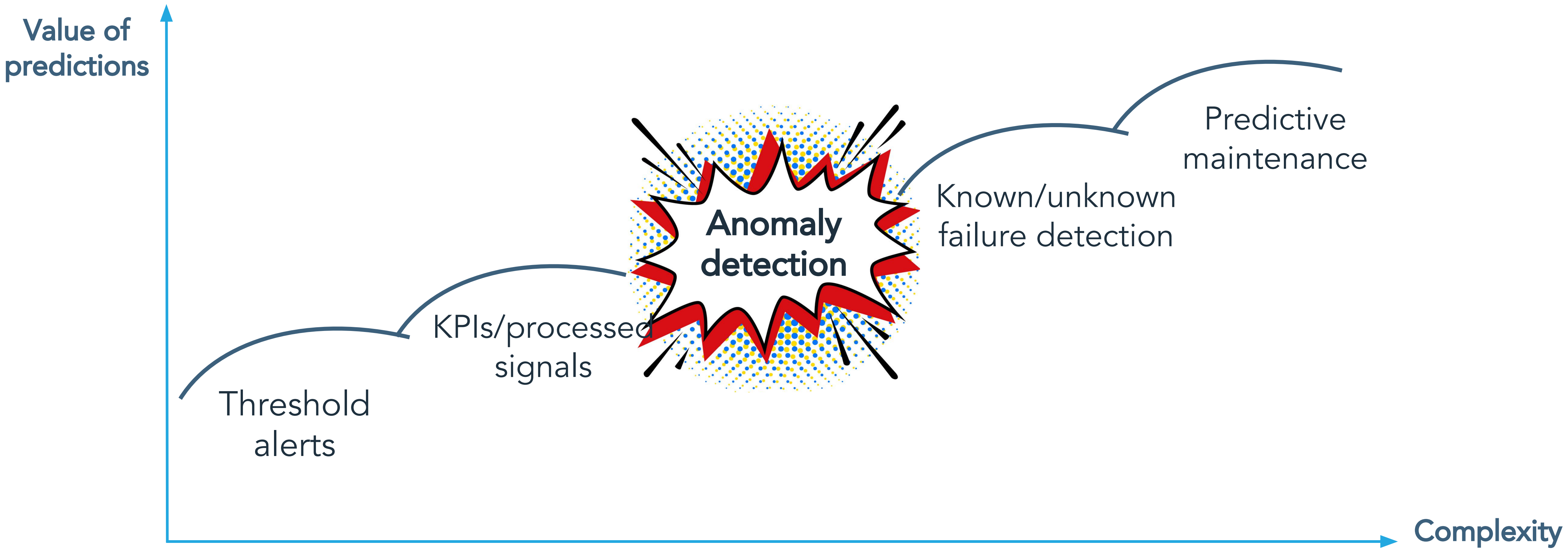
ARUNDO

Examples of anomaly detection



Copyright © 2016. Arundo Analytics. All rights reserved
ARUNDO

The stages of data-driven equipment monitoring

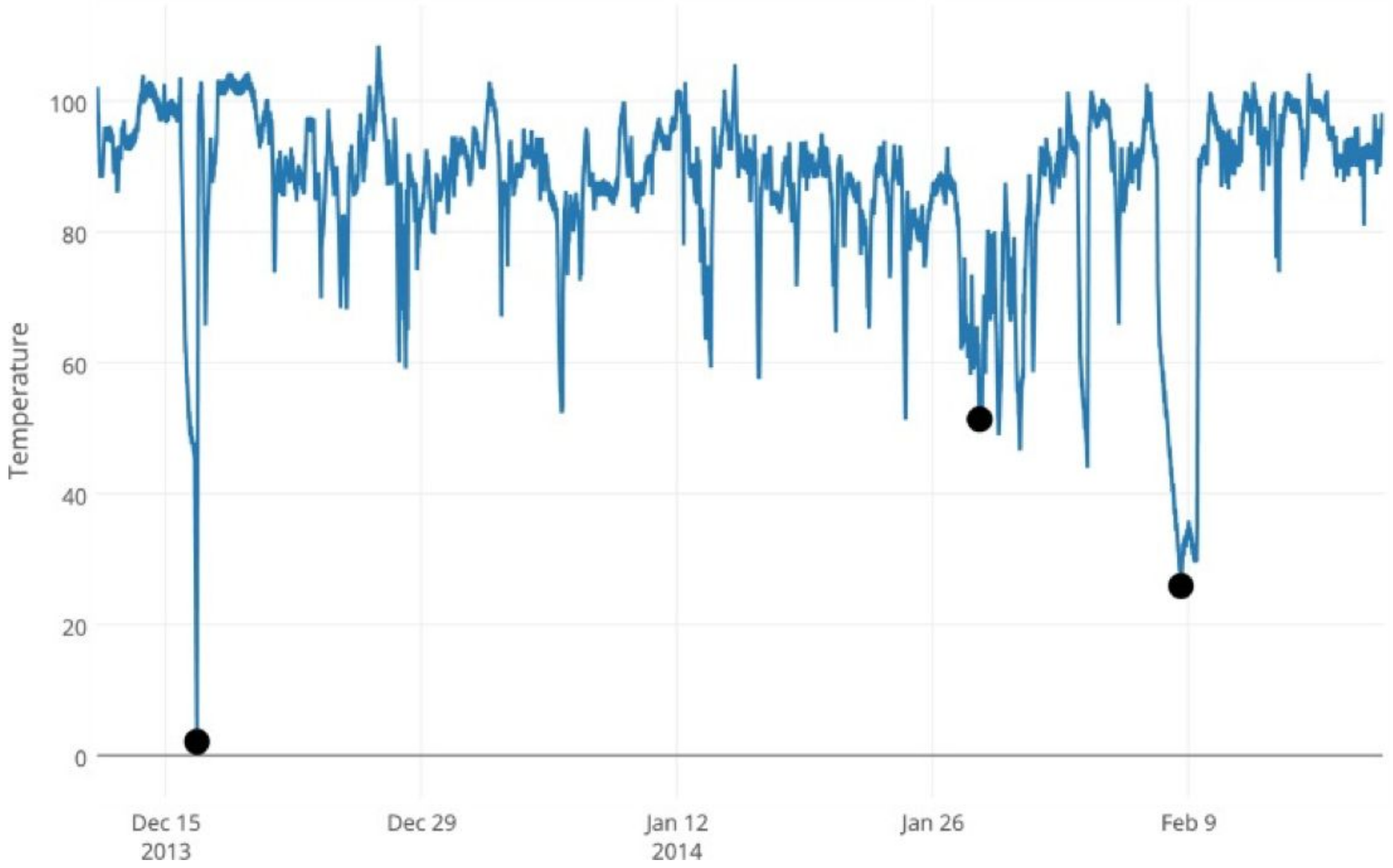


Copyright © 2018 Arundo Analytics, Inc. All rights reserved
ARUNDO

Anomaly detection in **equipment monitoring**

Previously unseen patterns can be a sign of:

- **misconfiguration**
- increasing mechanical **wear-out**
- **unforeseen** situations

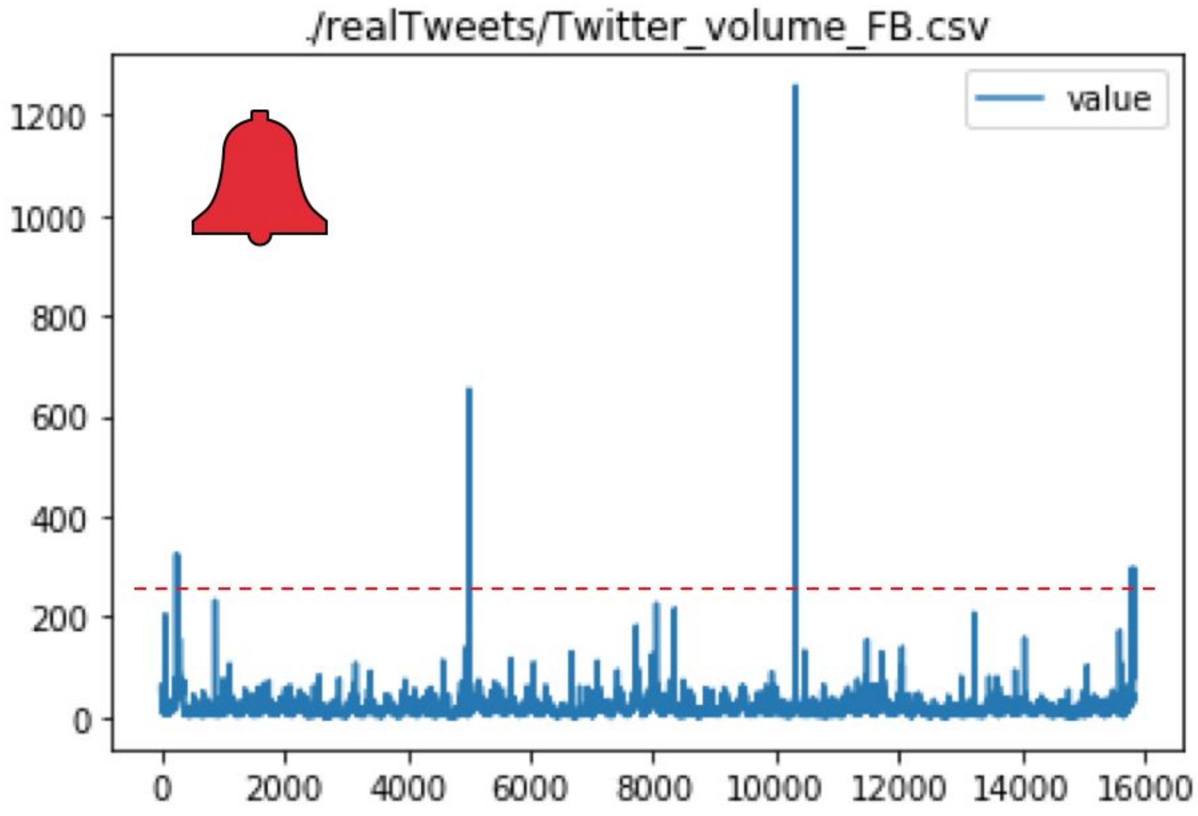


Copyright © 2018 Arundo Analytics, Inc. All rights reserved

ARUNDO

How can you detect anomalies?

- **Define a threshold for each sensor channel**
- **Raise a notification once a specified threshold is violated**



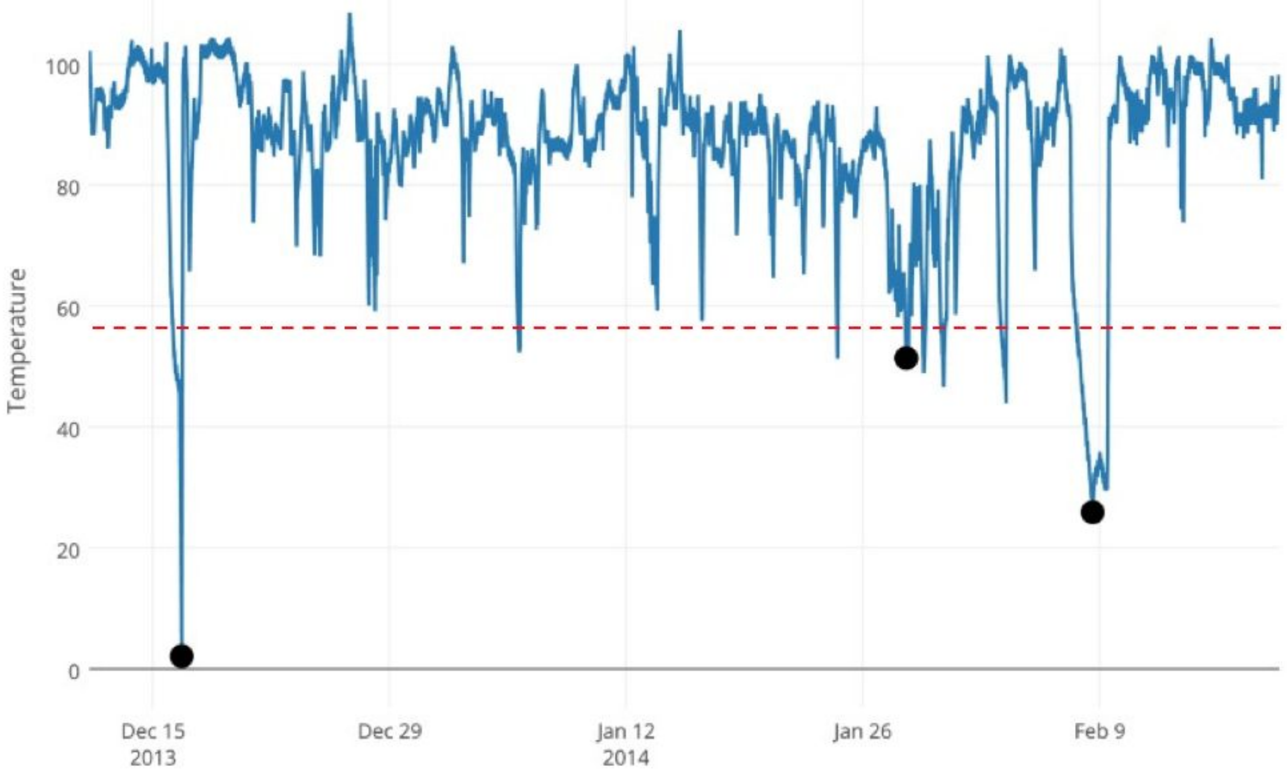
Copyright © 2018 Arundo Analytics, Inc. All rights reserved

ARUNDO

An oil rig can have upwards of 15.000 sensors
(the newest have more than 50.000 sensors!!!)

How can you detect anomalies?

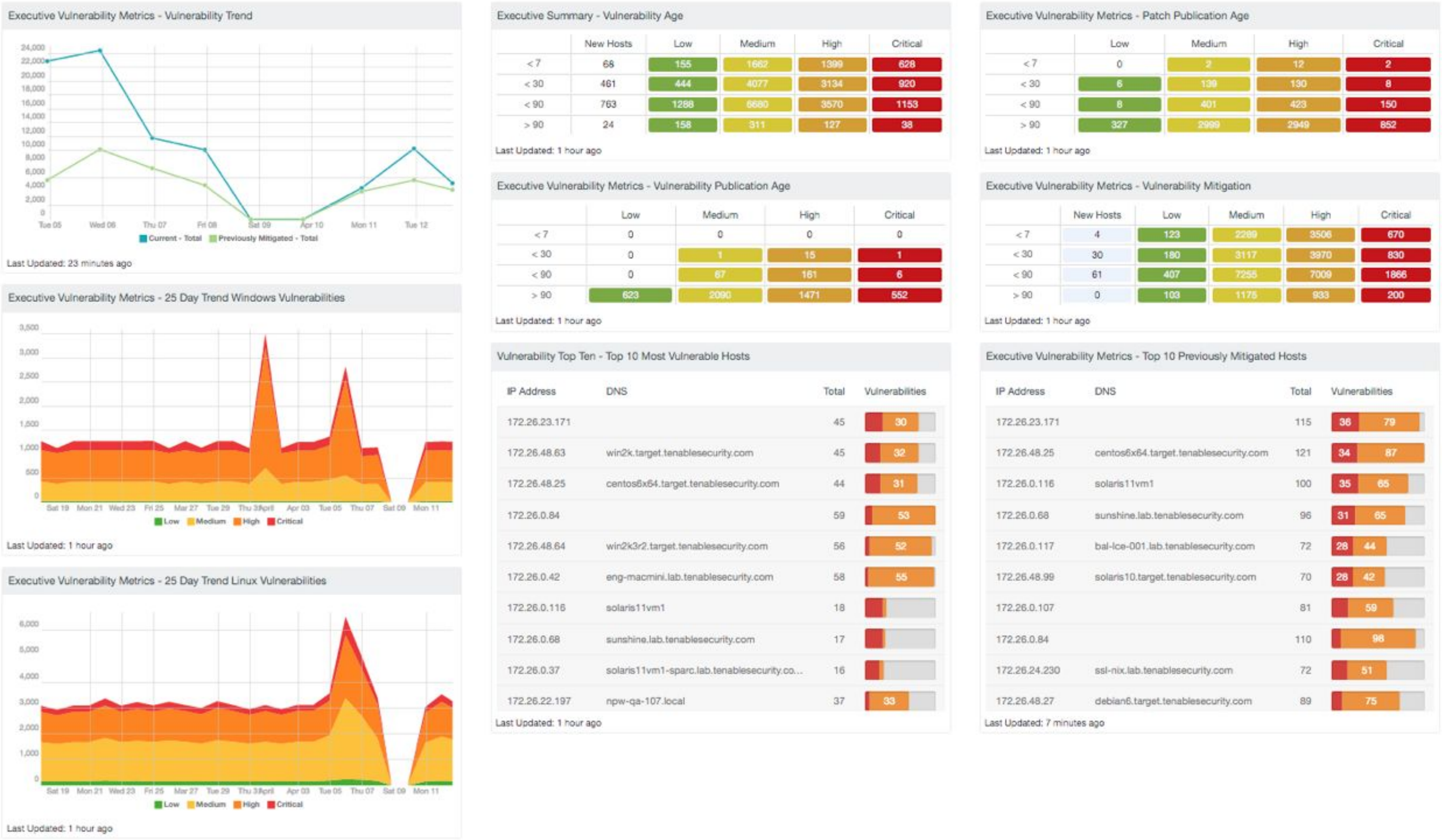
- **Causes many false alerts**
- **Does not take into account the joint characteristics of multiple channels**



Copyright © 2018 Arundo Analytics, Inc. All rights reserved

Manual analysis is immensely time-consuming and unreliable

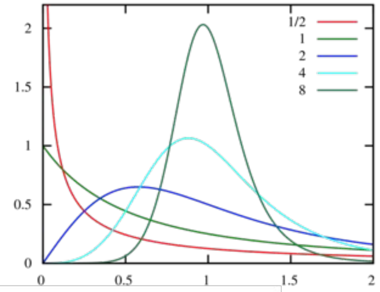
- **Massive** amount of multi-sensor data
- **Complex** systems
- **Rare** faults

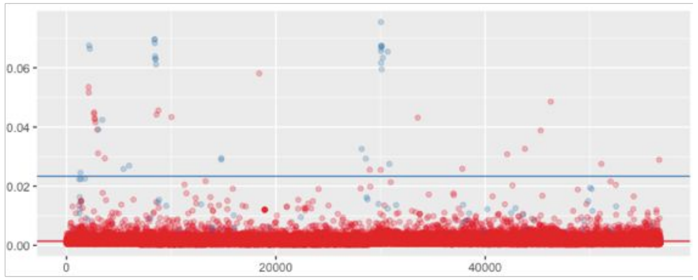


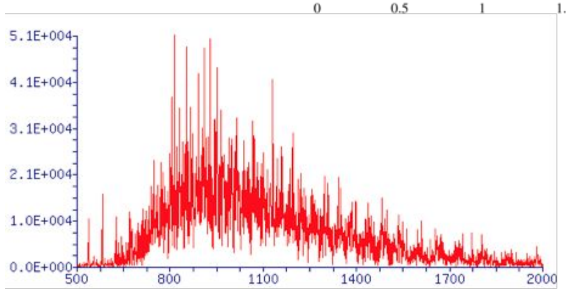
Copyright © 2018 Arundo Analytics, Inc. All rights reserved
ARUNDO

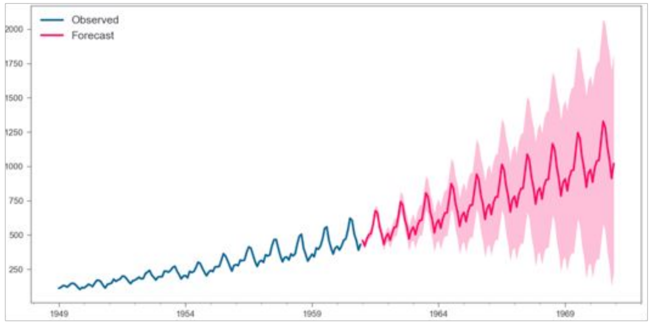
Multivariate anomaly detection

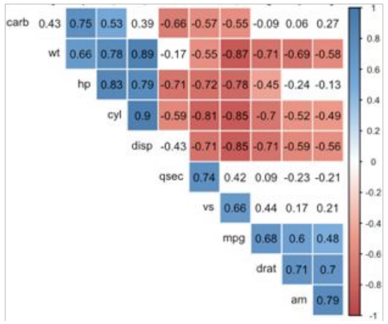
- No prior knowledge about anomalies
- No precise boundary
- Data often contain noise
- Normal behaviour keeps evolving
- Temporal dependencies
- Highly unbalanced classes
- High dimensionality and multimodal dependencies











Copyright © 2018 Arundo Analytics, Inc. All rights reserved

ARUNDO

Approaches

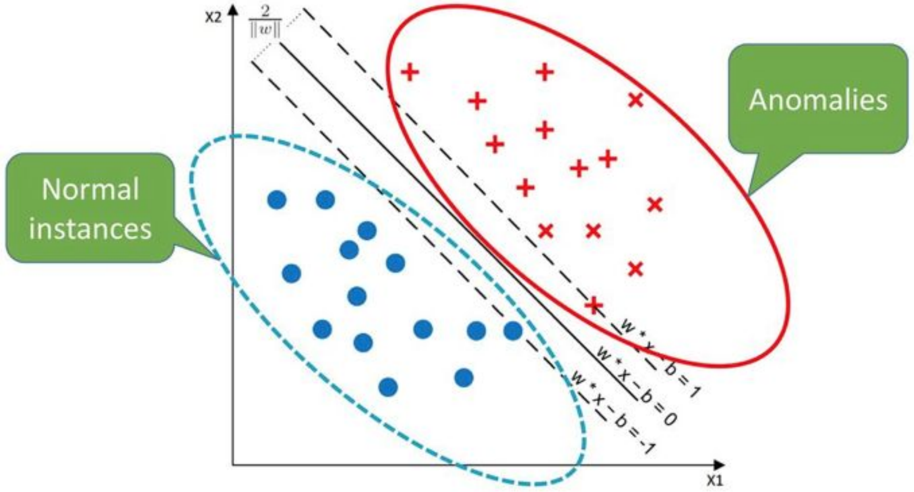
Data

Labeled data?
Predict Y from X?

Copyright © 2016. Arundo Analytics. All rights reserved

ARUNDO

Approaches



Data

Labeled data?
Predict Y from X?

yes

Supervised Learning

Develop predictive model based on both input and output data

Partition data based on labels

Classification

Regression

Copyright © 2016. Arundo Analytics. All rights reserved

ARUNDO

Approaches

Data

Labeled data?
Predict Y from X?

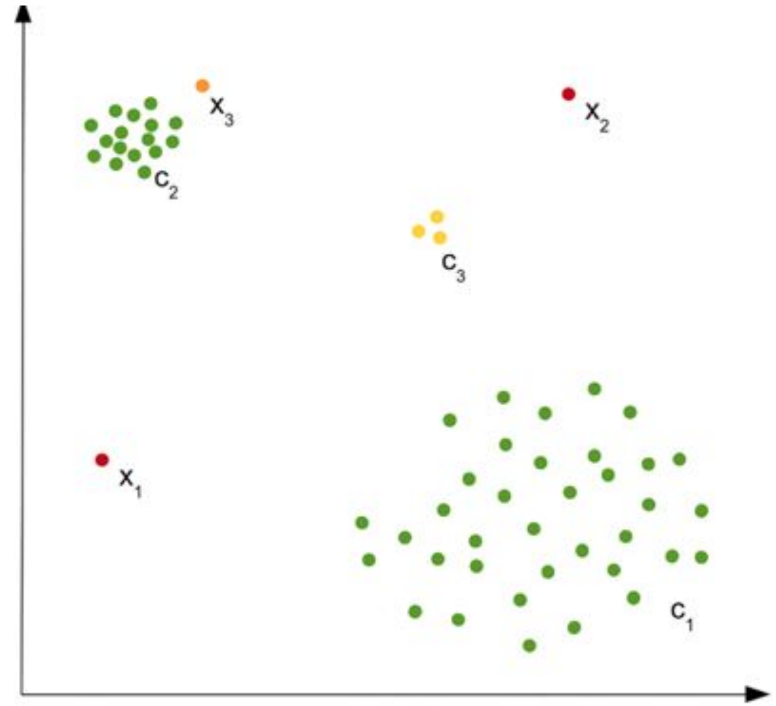
no

Unsupervised Learning

Group and interpret data based only on input data

Outliers in the distribution

Clustering



Copyright © 2016. Arundo Analytics. All rights reserved

ARUNDO

Real life example - Leakage on heat exchangers

Not enough failures to make a good model

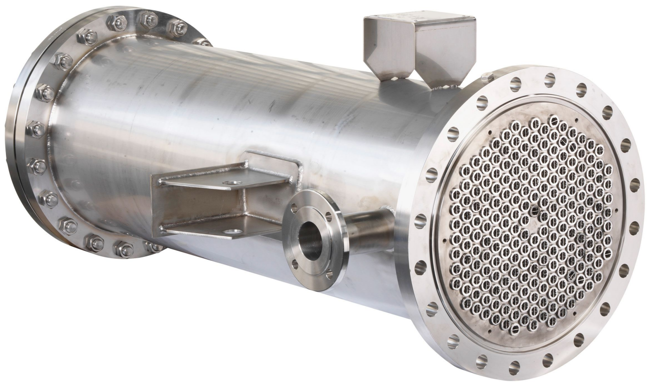
Unsupervised Learning

Group and interpret data based only on input data

What is the pattern of normal behaviour of the heat exchanger?

Sensor data+ maintenance logs

Training (simulation or historical) data available?



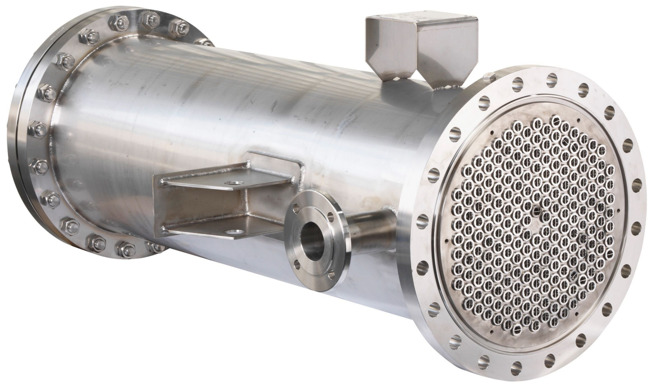
Copyright © 2016. Arundo Analytics. All rights reserved

ARUNDO

Real life example - Leakage on heat exchangers

Sensor data+ maintenance logs

Training (simulation or historical) data available?



Yes but not a lot of failures

Supervised Learning

Develop predictive model based on both input and output data

discrete

Has my heat exchanger sprung a leak?

continuous

What is the predicted performance of my heat exchanger?

Copyright © 2016. Arundo Analytics. All rights reserved

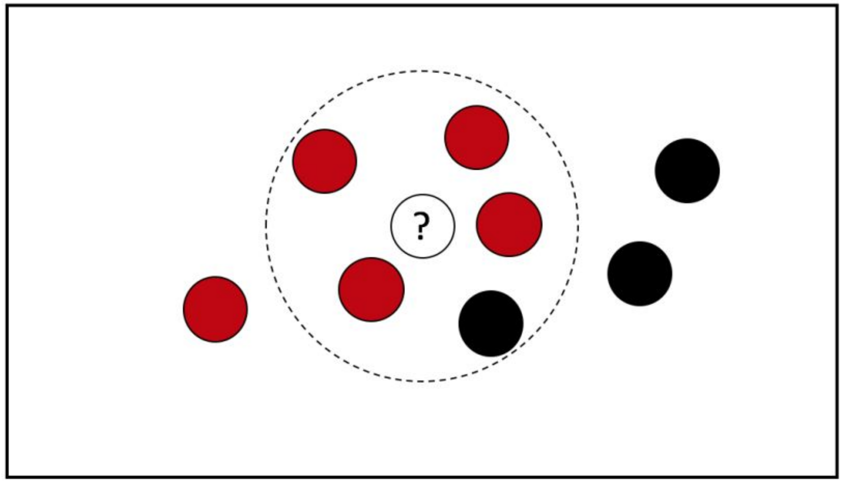
ARUNDO

Approaches

Supervised Methods	Unsupervised Methods
Random Forest	One Class SVM
Support Vector Classification	Isolation Forest
KNN Classifier	Elliptic Envelope
Logistic Regression	Local Outlier Factor
	Autoencoder

Copyright © 2018 Arundo Analytics, Inc. All rights reserved
ARUNDO

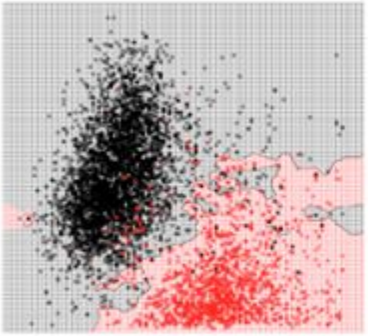
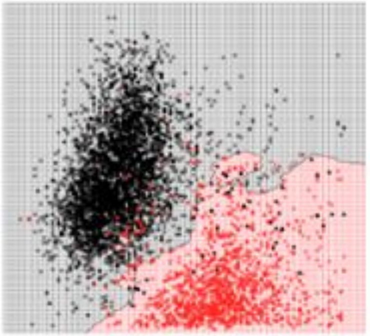
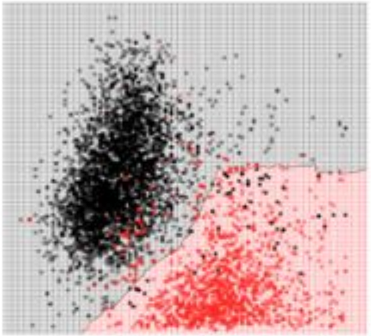
k-NN Classifier



_k_-NN classifies a data point based on how its neighbors are classified.

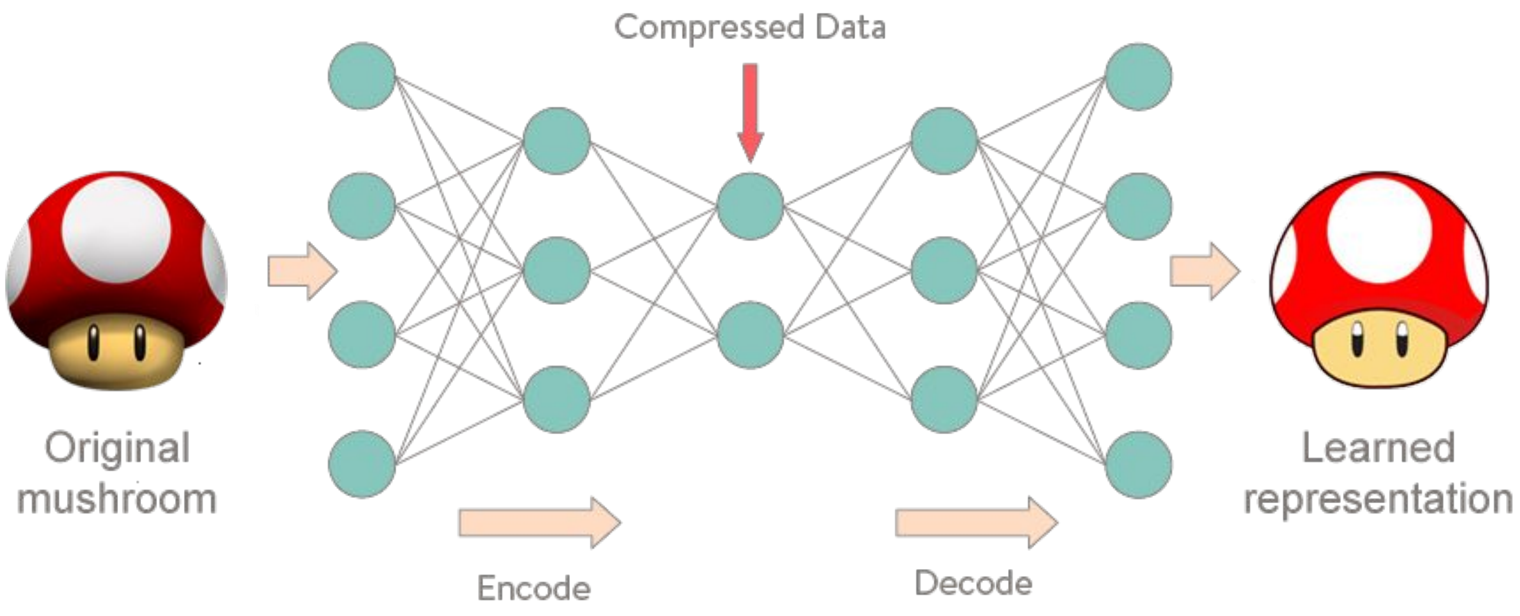
If a data point is surrounded by 4 red points and 1 black point, that data point is likely a red point by majority vote.

Tune for *k* - the number of nearest neighbors to include in the majority voting process

$k = 3$	$k = 17$	$k = 50$
		
98.2% accuracy	98.6% accuracy	97.8% accuracy
Overfit	Ideal fit	Underfit

Copyright © 2018 Arundo Analytics, Inc. All rights reserved

Autoencoders



The job of those models is to predict the input, given that same input.

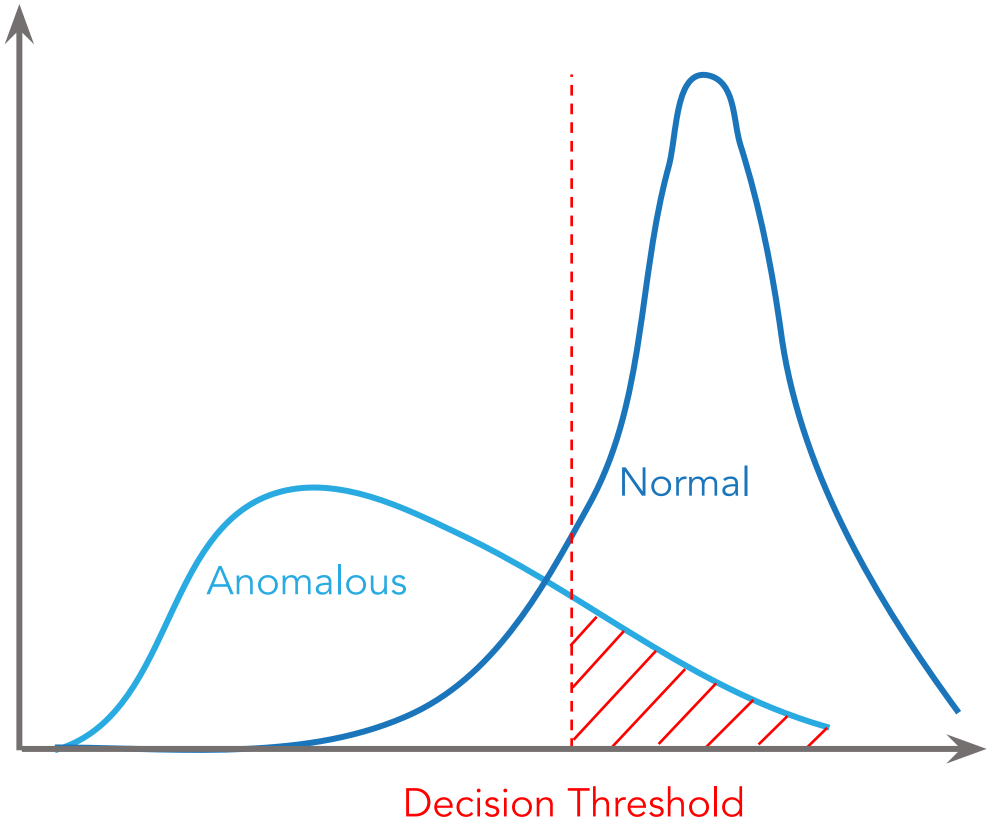
Learns a representation of the training data and recreates the input at the output layer.

Used for data compression and learning generative models from data.

We optimize the parameters of our Autoencoder model in such way that the reconstruction error is minimized.

Copyright © 2018 Arundo Analytics, Inc. All rights reserved
ARUNDO

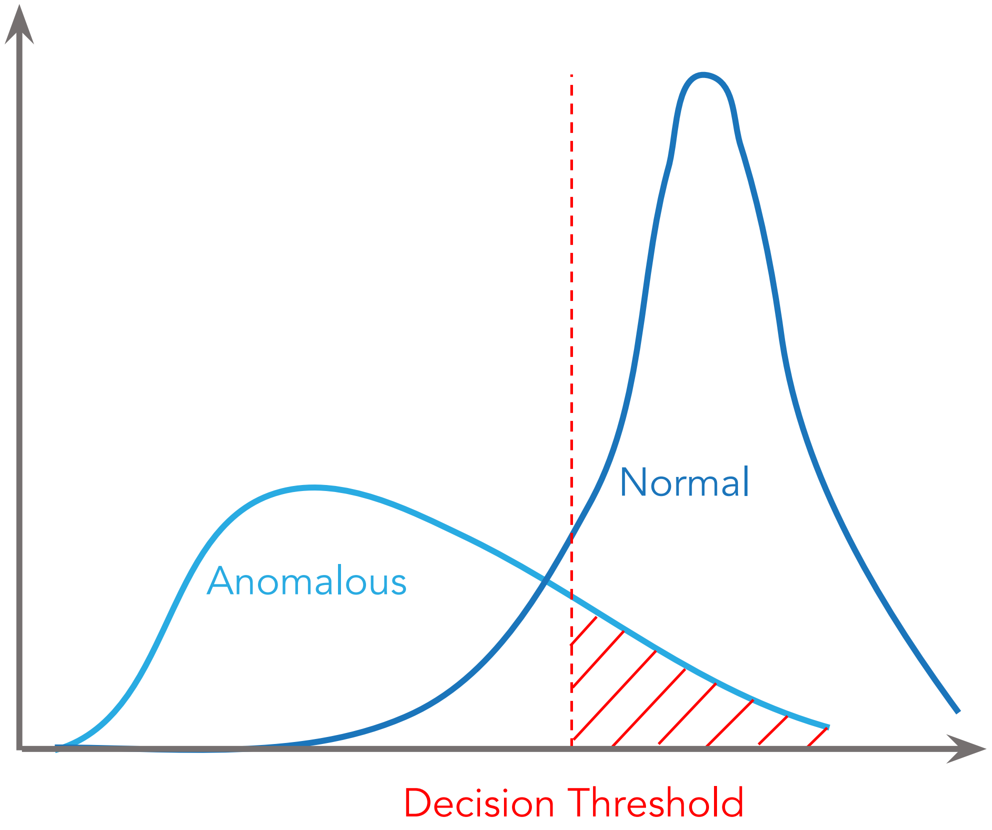
Machine Learning Predictions are not 100% Accurate



- False positives (ML predicts anomalous but system is normal) can and will occur
- False negatives (ML predicts normal but system is anomalous) can and will occur

Copyright © 2018 Arundo Analytics, Inc. All rights reserved

Machine Learning Predictions are not 100% Accurate



- Choose trained algorithms which minimize these effects as much as possible
- Identify any additional sources of data which may further help minimize these effects
- Evaluate the expected probabilities of each scenario using independent historical data
- Retrain the system if there is evidence that probabilities change significantly with time

Copyright © 2018 Arundo Analytics, Inc. All rights reserved
ARUNDO

Workshop details

Task

- Make a model to identify anomalies based on the dataset provided
- The model will be tested on a separate test dataset
- The best model will be chosen based on f1score

Practicalities

- Share e-mail via: goo.gl/forms/HY29cLwsqxiCJAMe2
- Log on to jupyter hub using your github account
    - appliedml-lausanne-2019.arundo.com

Notebooks

- Opening and looking at the data
    - ➢ 01-Model Development
- Model deployment
    - ➢ 02-Model Deployment
- Please add comments to explain your code where possible

Copyright © 2017 Arundo Analytics, Inc. All rights reserved
ARUNDO