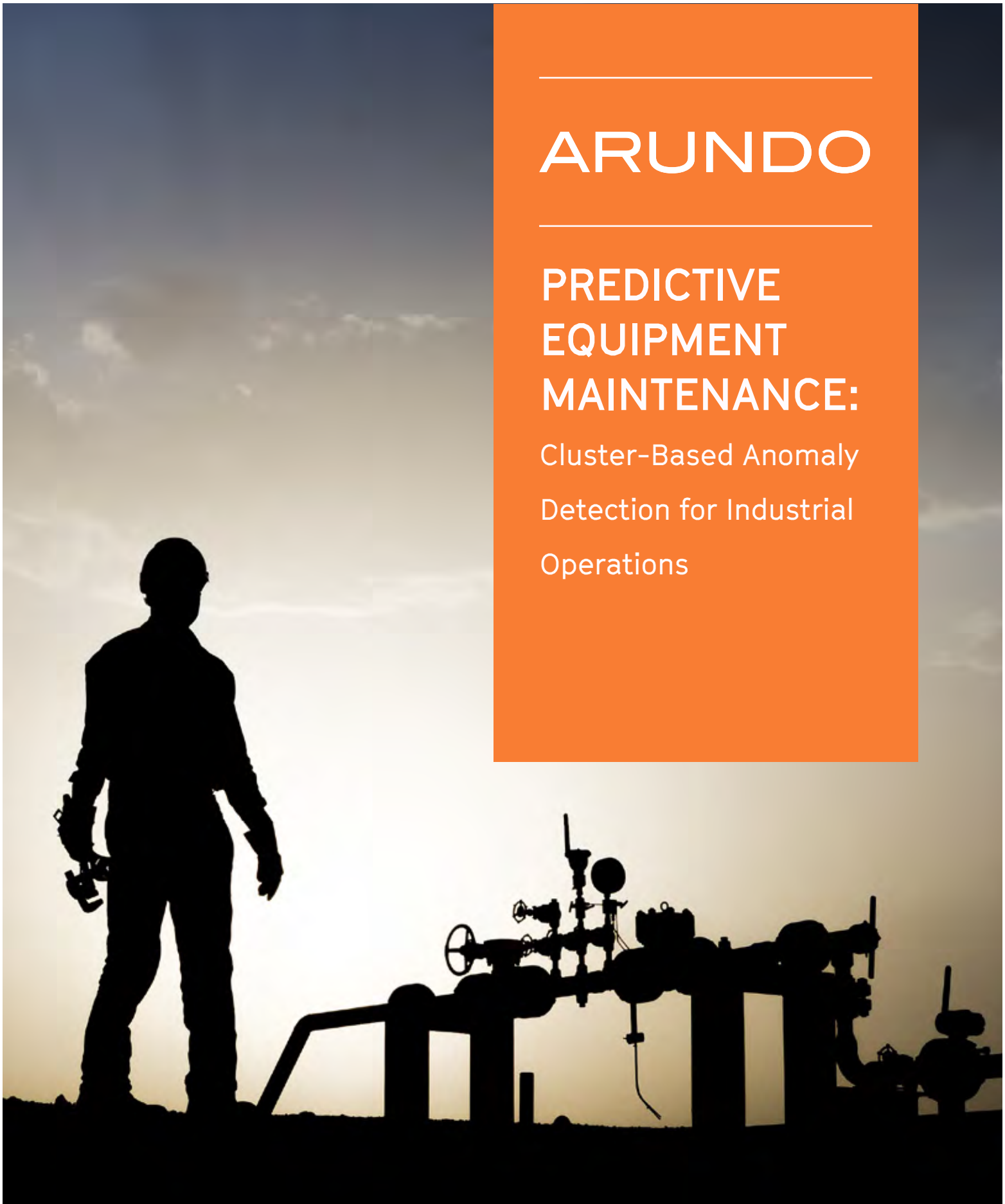
ARUNDO

PREDICTIVE EQUIPMENT MAINTENANCE:

Cluster-Based Anomaly Detection for Industrial Operations



Houston, USA
1600 Smith St Suite 3730
Houston, TX 77002

Oslo, Norway
Wergelandsveien 7
0167 Oslo

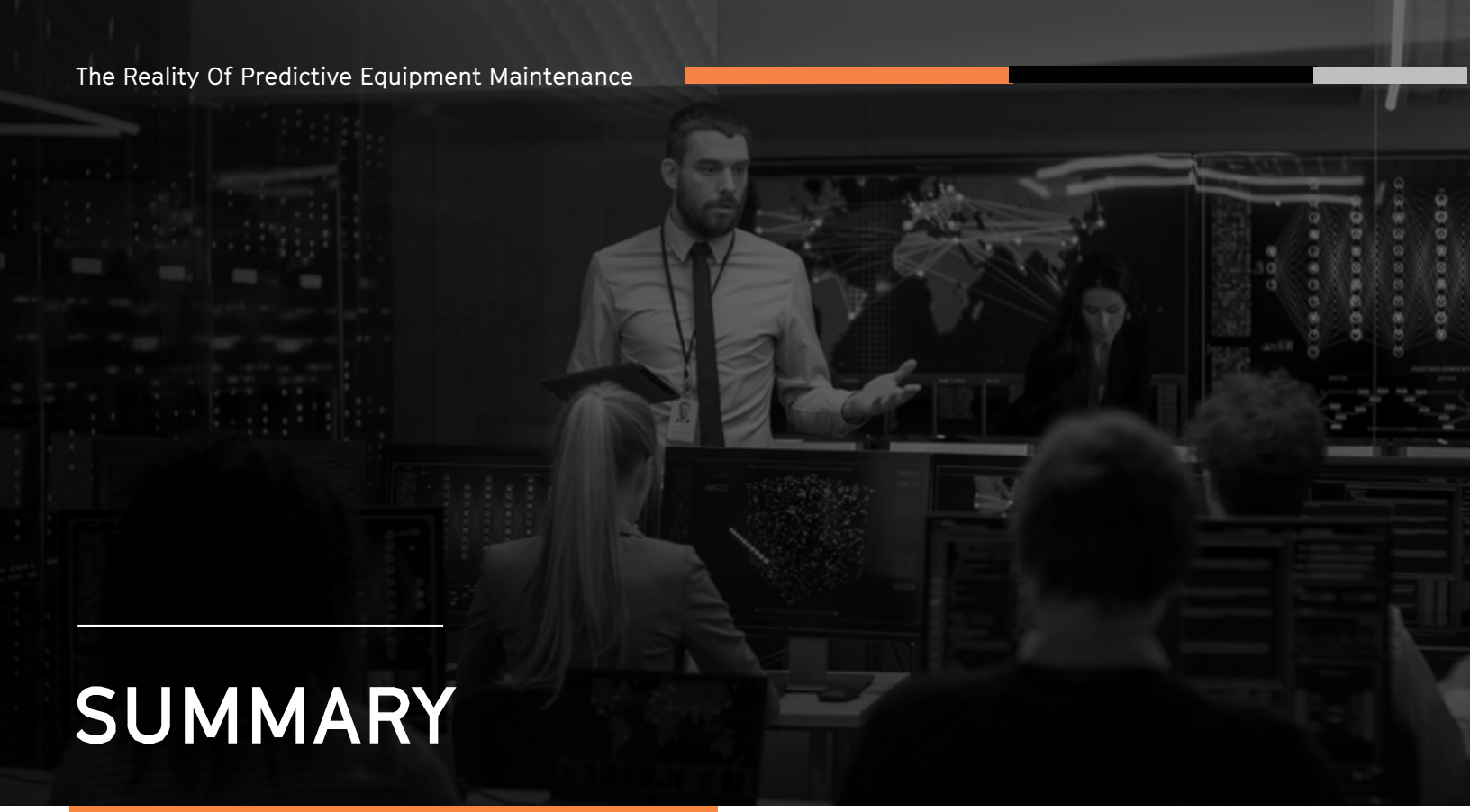
Palo Alto
444 Ramona Street
Palo Alto, CA 94301

CONTENTS

AUTHORS

Ellie Dobson, PhD, is VP Data Science at Arundo Analytics, a software company enabling advanced analytics in heavy industry. She works extensively on research and development of industrial data science applications, including anomaly detection for equipment monitoring. Ellie previously worked at Pivotal Software, The Mathworks, and CERN. Contact her at ellie.dobson@arundo.com

Amitav Misra is VP Marketing at Arundo Analytics. He works closely with industrial equipment manufacturers and major asset owners seeking equipment monitoring solutions. He previously worked at EnerNOC Inc. and McKinsey & Co. Contact him at amitav.misra@arundo.com



SUMMARY

This goal is more elusive than many business leaders may understand, however–it is rarely, if ever, an "out of the box" solution.

Achieving this goal requires a deeper understanding of the underlying data science and technology journey related to equipment monitoring and predictive analytics.

Heavy industrial companies are exploring the deployment of machine learning applications in order to improve revenue, reduce costs, and improve safety across their operations.
A common focus for these companies is to integrate predictive analytics - automated actions for certain highly likely outcomes - into their business processes. In particular, predictive maintenance for critical equipment is a key goal for many such industrial companies.

True predictive equipment maintenance involves automated notification of the specific mode of pending equipment failure: a message or alert to a human operator or control process that enables scheduled maintenance of the equipment at the lowest possible cost in terms of materials, labor hours, and equipment downtime.

In an ideal operation, the predictive maintenance systems tied into an automated scheduling system that can prioritize routes, materials, and work orders across a fleet of industrial equipment depending on predicted requirements.

In order for a machine learning model to produce such results, however, the model must understand (a) the universe of potential failure modes, (b) how patterns of sensor signals indicate specific failure modes, and (c) when such failures are likely to occur. Machine learning models need to learn. They need sufficient quantities of relevant data in order to make accurate predictions.

Despite the best efforts of data scientists and technologists, it is almost impossible to implement true predictive maintenance systems without significant historical operating data related to both sensor patterns and a large number of repeated, specific failures. This makes it difficult to assess the likelihood of failure of any specific piece of equipment.

Rather than attempting to develop predictive maintenance capabilities immediately, the most effective implementations of equipment monitoring take a “roadmap” approach, starting with streaming data capture and threshold-based alerts, and moving quickly into advanced anomaly detection as groundwork for the ultimate goal of a true predictive maintenance system.

THE CHANGING DATA LANDSCAPE IS TRANSFORMING INDUSTRIAL EQUIPMENT

Leading companies in heavy industries – operators and suppliers in energy, maritime, utilities, chemicals, and other capital-intensive operations – are reshaping their approach to operating performance in response to the convergence of several long-term technology trends:

- Data storage and processing capacity are effectively unlimited and almost free at the margin.
- Sensors continue to decline in cost and physical footprint.
- External data sources continue to proliferate.
- Sensor, device, and asset-level connectivity continue to improve in quality and cost.
- Machine learning tools and techniques are increasingly accessible.

This intersection of device connectivity, data storage, and compute capability is often referred to as the Internet of Things (IoT). For industrial companies, the ability to interact with physical equipment has never been greater. Companies now have the ability to access and analyze a previously unimaginable amount of data, arriving almost constantly from a variety of sources, and to make meaningful business decisions from this data continuously.

Industry leaders are taking advantage of this ability to increase revenue, decrease costs, and to create new business models.

However, most companies are just beginning their digital transformation – they typically access, analyze, and make decisions based on just a tiny fraction of the potential data generated from their assets and equipment.

Table 1. The Changing Data Landscape for Industrial Businesses

	Legacy system design constraints	Industrial Internet of Things potential system capabilities
Size of datasets generated	Megabytes to gigabytes	Terabytes to petabytes
Timing of data availability	Weekly or daily batches	Near real-time to real-time streaming
Typical analytical approach	Periodic asset-level optimization with linear programming models	Continuous system-wide predictions and optimization with machine learning models
Decision-making integration	Periodic reports; dashboards	Real-time alerts; continuous process feedback

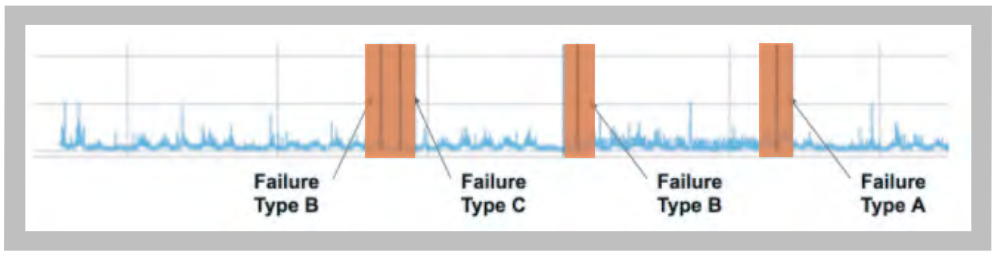
*Source: Arundo interviews and analysis

Every industrial company must come to terms with the emerging data landscape created by the Industrial Internet of Things (IIoT). Already, many asset owners and operators are “sensoring up” their physical operations – even before nalizing the new business strategies, operating processes, and software tools required to realize value from new digital assets and data streams.

PREDICTIVE EQUIPMENT MAINTENANCE IN THEORY AND PRACTICE

In true predictive maintenance applications, a piece of equipment, such as a pump, compressor, or heat exchanger, alerts human operators or other systems about a specific mode of impending failure in time for an intervention that avoids unnecessary downtime or expense. In order to fully achieve this goal, a machine learning model must be trained on fully representative historical data, with all failure events accurately labelled-i.e., there must be many examples of all possible failure events labelled on historical sensor data. However, such datasets are rarely available in actual industrial operations.

Figure 1. Example historical sensor data with labelled failure events

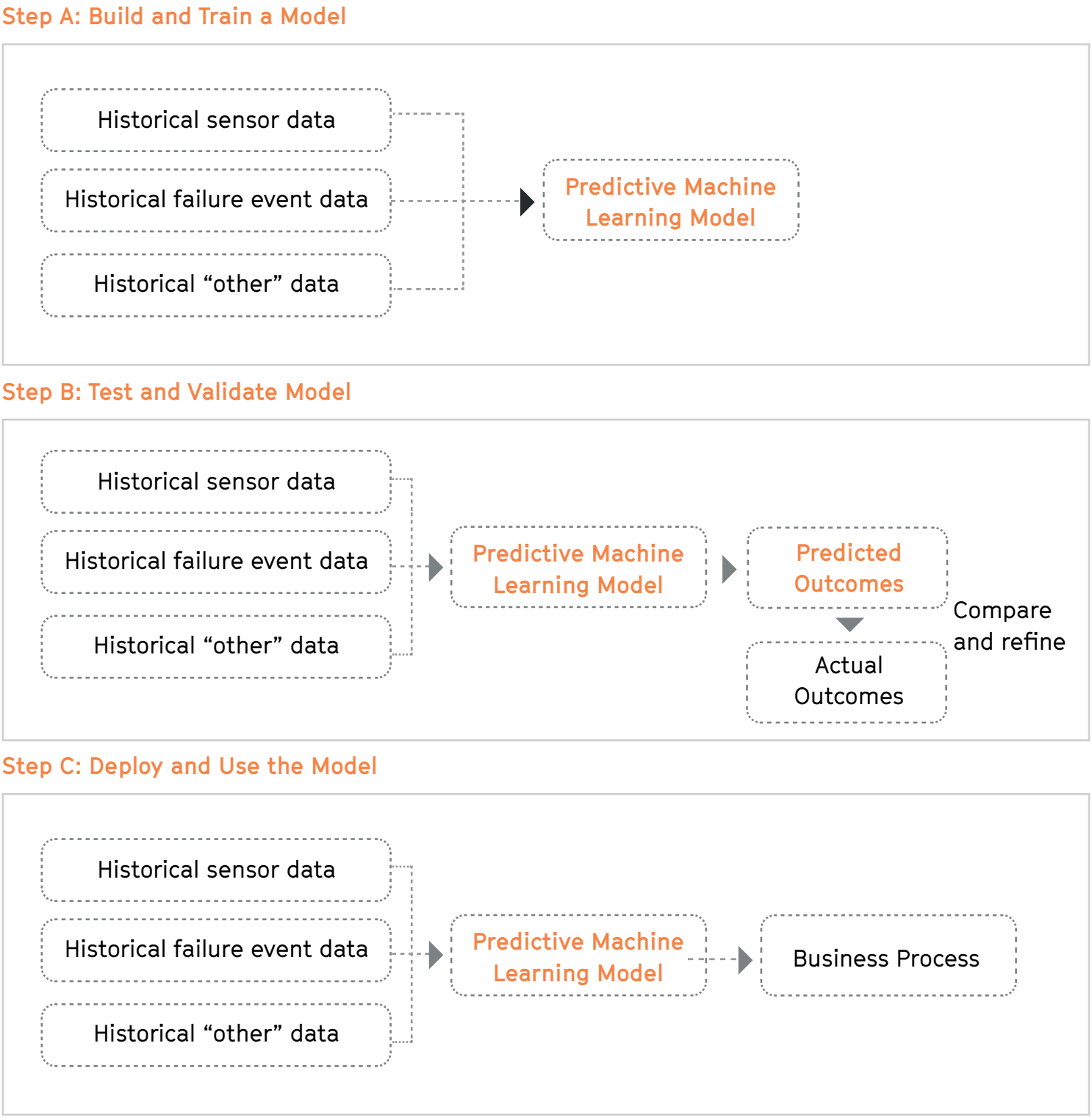


For most field-installed equipment, data is commonly available in one of the following scenarios:

1. There is no historical sensor data, but new sensor data is streaming from the equipment.
2. There is historical sensor data, but it is not labelled with historical failure events
3. There is historical sensor data, and there is historical failure data, and with manual or automated pre-processing, the datasets can be joined to get labelled sensor data - however the failure events are relatively rare, and don't fully represent the entire universe of potential failures in sufficient number for automated predictions.be joined to get labelled sensor data rare, and don't fully represent the entire universe of potential failures in sufficient number for automated predictions.

In practice, due to the nature of developing and testing machine learning models, these common data situations make the immediate application of true predictive maintenance systems quite rare for most heavy industrial equipment.

Figure 2. Building and using a machine learning model



The quality and accuracy of machine learning model outputs are largely driven by the availability of large amounts of historical failure data. Without a sufficient number of historical failures, even the most sophisticated machine learning techniques are often futile in predicting failures with better accuracy than existing simulations or physics-based estimates. Deep learning, for instance, typically requires tens of thousands of observations to be effective.

FROM SINGLE-SENSOR THRESHOLD ALERTING TO CLUSTER-BASED ANOMALY DETECTION

Once equipment is properly instrumented, a common initial approach is to stream, capture, and visualize sensor data. This may be combined with threshold-based alerts for individual sensor values (for instance, if temperature or vibration fall above or below specified levels, an automated notification is sent to certain users.)

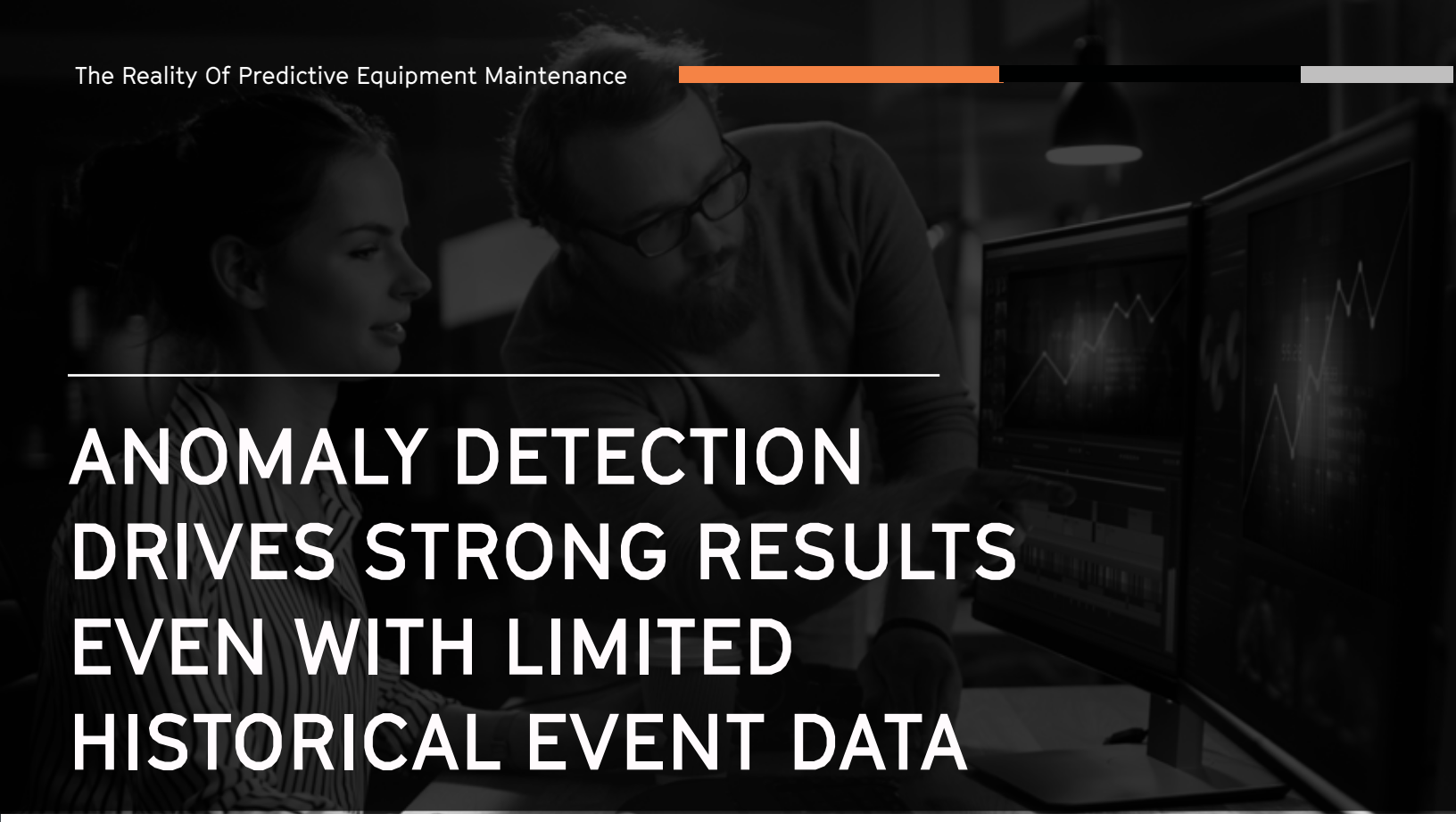
However, industrial equipment typically exhibits a range of complex behavior, which may lead to challenges with threshold-based alerts. Often, the equipment may have multiple operational modes. Even normal operations may display performance along an operational curve. Traditional single-sensor alerting systems have limited utility under these circumstances.

They may raise unnecessary alarms when in fact the equipment is simply in a corner condition of normal operations, or perhaps a different operational mode such as ramp-up. False alarms are a significant challenge for equipment monitoring systems in heavy industry, as they often result in operators ignoring alarms altogether, and failing to act in advance of major equipment failure.

As noted above, true predictive maintenance is not immediately applicable for most equipment, due to the paucity of relevant data. However, anomaly detection, a related type of machine learning-based system for equipment analytics, can be used with significantly smaller datasets.

Arundo developed its anomaly detection system based on many years of historical data from numerous pieces of industrial equipment. This approach relies on a set of algorithms belonging to the unsupervised learning technique called clustering, which learn from historical data and build up a much richer and complex view of system behavior across groups of sensors. The trained model can then raise alarms when previously unseen or known failure behaviors arise in the system.

This anomaly detection approach accounts for multiple complex operational modes. It also enables the combination of multiple sensors into a single measure (a “virtual sensor”) of the health of the system. This results in the system raising fewer false alarms, while at the same time providing a more accurate prediction of anomalous behavior.



ANOMALY DETECTION DRIVES STRONG RESULTS EVEN WITH LIMITED HISTORICAL EVENT DATA

At a major national oil company, we used anomaly detection to identify potential compressor failures with adequate operational lead time for inspection, maintenance, or repair. The anomaly detection approach is now integrated into the company's equipment monitoring system.

A hydroelectric power plant exhibited complex operating behaviors. These included varying operating conditions in ramp-up, normal operation, and ramp-down modes. We deployed a density-based clustering model for anomaly detection that provides higher confidence estimates of potential failure than any physics-based model, engineering simulation, or equipment manufacturer estimates.

At a major offshore oil & gas producer, we provided anomaly detection as an integrated equipment condition monitoring application for heat exchangers and compressors. This is delivered through an application with alert functionality and ability for equipment specialists to access all information related to failures (sensor data, technical drawings, etc.) in one integrated view for rapid diagnosis to take preventive actions. We are working with the company to deploy similar applications for additional equipment categories.

In aggregate, these applications of anomaly detection have the potential to save tens of millions of dollars across the three companies on this equipment alone.

IMPLICATIONS FOR INDUSTRIAL EQUIPMENT OWNERS AND SUPPLIERS

Industrial equipment is moving away from decades – sometimes centuries – of incremental mechanical improvements. Today, fundamental data-driven insights are changing the way industrial equipment is built, sold, and operated.

Owners, operators, and suppliers of such equipment must adjust to the new data landscape. The ultimate goal of true predictive maintenance may be a dificult short-term objective. However, equipment condition & performance monitoring solutions that integrate anomaly detection techniques across multiple sensors present a critical first step.

With such systems in place, industrial equipment can join the Industrial Internet of Things and the digital transformations that are changing every aspect of heavy industrial operations.