

DATA ANALYTICS FOR IoT

Business Transformation
Through Connected Products

WHITE PAPER

RELEASED: FEBRUARY 2016

Author: Mark Benson & Jay Carlson

Revision: A

TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. CLOSING THE GAP TO DECISION.....	1
2.1 Data Analysis and Analytics for IoT.....	1
2.2 Benefits of Data Analysis.....	2
2.3 Data Collection Requirements.....	2
3. ANALYTICS MATURITY.....	3
3.1 Descriptive Analytics: What happened?.....	4
3.2 Diagnostic Analytics: Why did it happen?.....	4
3.3 Predictive Analytics: What will happen?.....	5
3.3.1 Cluster Analysis.....	5
3.3.2 Classification.....	5
3.3.3 Random Forests.....	9
3.4 Prescriptive Analytics: What should be done?.....	9
3.4.1 Decision Support vs. Decision Automation.....	9
4. TURNING INSIGHT INTO ACTION.....	10

“DATA WITHOUT
CONTEXT IS
MEANINGLESS.”

1. INTRODUCTION

The era of big data is upon us. With the advent of the Internet of Things (IoT), companies have access to more information than ever before. Employees sit in Internet-connected chairs in rooms with remotely monitored HVAC systems. Industrial buildings, built by cranes with pumps whose performance can be seen on an iPad, feature Internet-connected pest-control traps that send text messages to sanitation workers when rodents have been caught. Employees go home to Internet-connected garage doors, lights, and thermostats that can all be controlled from a smartphone.

Although a staggering amount of information is gathered every day from devices just like these, that data is not necessarily being used effectively by the companies that have access to it. As Tim Hartford of the Financial Times notes, “Big data has arrived, but big insights have not.”¹ Useful data analysis requires much more than the simple collection and summary of data. Companies must have a long-term analytics strategy in place to provide significant, actionable insights that will fuel their business transformation from a product company to a connected product company. That strategy must begin with a solid foundation of connected devices and include the purposeful creation of a data pipeline, from developing processes for data gathering to leveraging algorithmic predictions to inform business decisions.

This white paper provides an overview of analytics within the context of connected products and identifies the stages of analytics maturity. Along the way, it introduces machine learning and the prospect of automating device behavior. Finally, it concludes with

recommendations about how enterprises building connected products can realize the full potential of IoT by turning data into action that produces business results.

2. CLOSING THE GAP TO DECISION

2.1 DATA ANALYSIS AND ANALYTICS FOR IoT

Closely related to statistics, data analysis is the process of gleaning new insights out of data sets with the purpose of more intimately understanding the data and the behavior it explains. Historically, most data analysis has been done manually with languages like R, Python, or SAS. Technological advancements and the explosion of IoT have not necessarily relegated manual data analysis completely to the past, but it is now often more prudent to encode many of these advanced analytics into machines.

Much like analytics in other more traditional contexts, analytics for IoT is one of many tools that enables data analysis. However, real-time sensor data made available by IoT-enabled devices provides new opportunities for on-the-go analysis and programmed decision making, which results in a type of visibility into products that was previously unattainable.

At their most basic level, data analysis platforms can be easily integrated with existing databases of meta-data and sensor data. Another level up, real-time data is complemented with real-time stream analytics, which can trigger pre-programmed decisions at the same rate that data is being gathered. Further, in

¹<http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html>

situations where Internet connectivity is not reliable or where the amount of data being passed to the cloud is purposefully limited, analytics code can be pushed directly to a device or gateway in order to ensure proper operation even in the face of an unreliable or missing network connection.

2.2 BENEFITS OF DATA ANALYSIS

Data analysis allows companies to operationalize business insights and device behaviors quickly and easily. As data is utilized more and more effectively, the distance human input has to traverse in order to make a decision can be greatly reduced. The ability to make empirical and grounded, what was once hunch and assumption, is the hallmark of data analysis. In the increasingly competitive landscape of connected consumer and industrial products, the maturity of a data analysis program may mean the difference between merely surviving in the connected market and fully transforming into an industry-leading connected product company.

A mature data analysis program can provide new avenues to reduce costs or monetize a connected product, shedding light on customer segmentation, predicting maintenance needs, and preventing unintended and costly downtime.

As analytics strategies increase in sophistication from descriptive through prescriptive, the human input required to turn information into action can be dramatically reduced, minimizing costs and improving efficiency, results, and customer experience.

These outcomes can only be realized when based on solid data collection standards combined with the stages of analytics maturity that comprise a

full-fledged IoT analytics solution.

2.3 DATA COLLECTION REQUIREMENTS

A standardized data collection process is comprised of three parts:

1. Pinpointing the questions that need to be answered with data analysis.
2. Identifying what data is needed to support the answers to those questions.
3. Gathering the correct type and amount of that data.

Although the value proposition of a robust data analytics program can be astounding, it is useless if the proper data is not gathered. Even the most sophisticated analysis cannot synthesize or replace foundational input data. Once unanswered questions have been identified, it is important to determine the metadata, sensor data, and data resolution that best fits a particular situation, which can be a process in and of itself.

Exploratory techniques such as data mining take in massive amounts of data and find correlations that might otherwise go unseen. By feeding in sample data that includes a number of outcomes that are ideal to predict, data mining can make the process of determining what variables are truly indicative of device behavior as simple as running a program. However, this pursuit of predictors and causal relationships in data oftentimes merely illuminates what is already known. In-house expertise is absolutely irreplaceable in determining how products will behave. For instance, pump designers do not need extensive data analysis to tell them that vibration, temperature, and pressure are key metrics when

predicting pump failure. In many cases, an engineer's intimate understanding of a product will point data analysis precisely in the direction it needs to go.

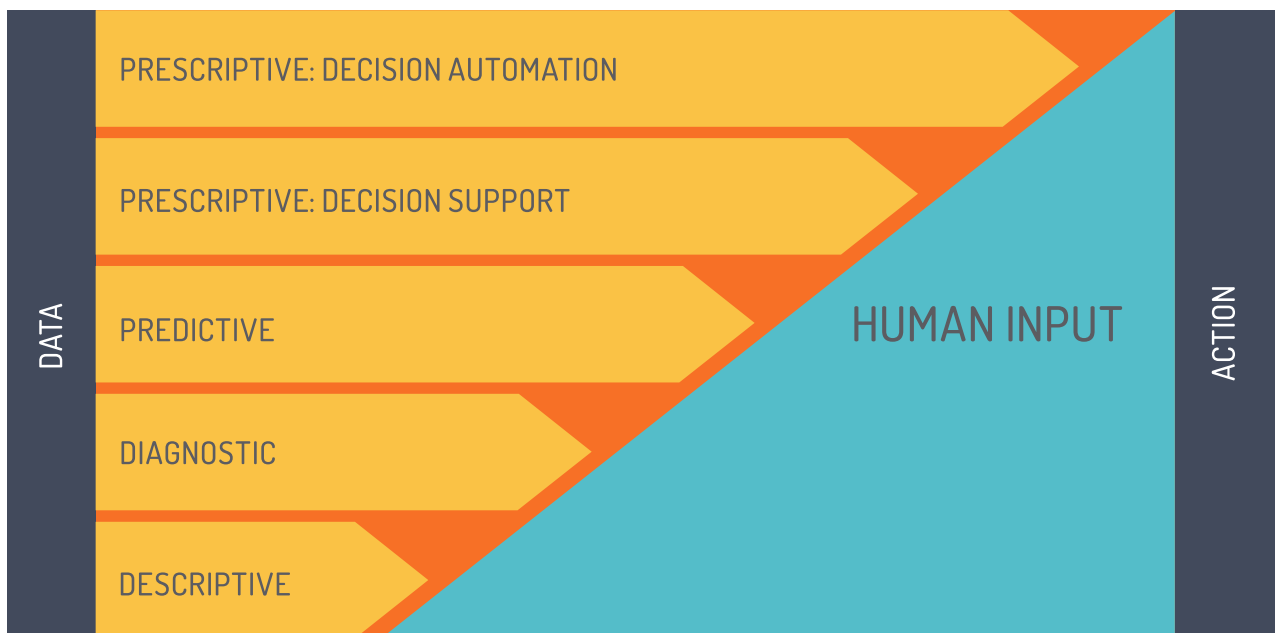
Once the necessary data has been defined, that data must be collected in a reliable and consistent manner. It is important to keep in mind that the amount of data that needs to be captured will vary across applications. For instance, some situations will call for data to be sent from a single sensor every fraction of a second. In other situations, sensor data need only have minute resolution. Collaboration between product and IoT experts will yield appropriate data resolution standards to ensure that cloud connectivity is being used efficiently, while still capturing impactful results.

3. ANALYTICS MATURITY

Once data collection standards have been defined, companies can begin progressing through the stages of analytics maturity to decrease the level of human input required to take data and turn it into action. The stages of analytics maturity are shown in Figure 3.1 below. The following is a discussion of these stages of maturity, including descriptive, diagnostic, predictive, and prescriptive analytics.

FIGURE 3.1: STAGES OF ANALYTICS MATURITY

As analytics strategies increase in sophistication from descriptive through prescriptive, the human input required to turn information into action can be dramatically reduced, minimizing costs and improving efficiency, results, and customer experience.



3.1 DESCRIPTIVE ANALYTICS: WHAT HAPPENED?

The first step in data analytics maturity is describing what is happening with the devices in a connected fleet. What is the average pressure inside a device? What is the device's maximum runtime each day? What is the greatest period of time that a device is at its highest pressure?

The measurements and visualizations typically seen in descriptive analytics include moving average, standard deviation, histograms, and quartiles. Descriptive analytics do not really dive into the causes of device behavior or even look much further than the facts themselves. However, descriptive analytics are the building blocks of a mature analysis program; even at this level, it is possible to see how descriptive analysis leads to action.

Imagine a pump in an industrial application that has an average pressure of 600 PSI and never exceeds 650 PSI for more than a few seconds. Descriptive analytics can help identify when a pump is operating at pressure levels outside this normal range. For example, a technician with this type of descriptive information is much more likely to notice that a pump has been running at 700 PSI for several minutes and identify this as a potential problem. In this way, descriptive analytics enables troublesome behavior to be addressed immediately rather than letting it run its course, helping avoid device failure and potentially thousands or millions of dollars in damage and lost productivity.

3.2 DIAGNOSTIC ANALYTICS: WHY DID IT HAPPEN?

Upon digging deeper into the data provided by descriptive analytics, it is possible to see that failure or unexpected behavior of a device is often linked to an anomaly in its behavior before the event. What state was the device in when the fault occurred? What may have contributed to the fault? Is different data needed to make a more accurate assessment of why the device failed?

Diagnostic analytics tell us what previous device behavior likely caused or indicated its behavior. This can be as simple as identifying that an abnormally high value for one metric, or a combination of metrics, typically preceded a specific device behavior.

Continuing with the industrial pump example above, vibration is one of the most prominent causes of pump failure. Vibration can be caused when a bearing fails, or when a pump is subjected to a shock that it was not built to withstand. Diagnostic analytics could easily link the descriptive statistics of an increase in vibration to the eventual failure of a pump, as long as pump vibration is being monitored. This type of diagnostics would allow technicians to take actions to limit vibration going forward or to request that higher vibration thresholds be incorporated into future pump designs.

If the cause of a device's failure cannot be diagnosed from descriptive statistics, this is a sign that more information needs to be collected, which could mean obtaining additional metrics or sampling data with higher granularity.

3.3 PREDICTIVE ANALYTICS: WHAT WILL HAPPEN?

Once diagnostic analytics have been used to identify key causal relationships between past device behaviors and events, it is possible to leverage those retroactive insights to begin projecting probable device behavior into the future.

The benefits of knowing what a device fleet or external entity is going to do are far-reaching. It allows technicians to schedule proactive maintenance to reduce unexpected downtime, enables organizations to save thousands or millions of dollars in damage caused by mal-performing equipment, and eliminates the potential for degraded brand trust caused by unexpected device behavior.

Making the leap from predicting future device behavior to developing a system that automatically takes action based on those predictions can be complex and often involves extensive algorithms that force device behavior.

Machine learning is a system in which computers apply statistical learning techniques to automatically identify patterns in data, modify those patterns as more data is gathered, and make highly accurate predictions. Machine learning algorithms are trained with control and anomalous data to recognize normal and abnormal device behavior. They can also be retaught when conditions for failure change if, for example, a device upgrade is made.

Machine learning is an incredibly powerful tool for predictive analytics in IoT applications and has massive efficiency and performance ramifications for manufacturers. With mature machine learning analytics in place, manufacturers can limit downtime

and liabilities as their system predicts and self-diagnoses failure and adapts to new device conditions to operate effectively regardless of context. Although there are many machine learning techniques in use today, the discussion below focuses on a few key concepts, including cluster analysis, classification, and random forests.

3.3.1 CLUSTER ANALYSIS

Cluster analysis is a huge component of data analysis for IoT. It involves predicting device behavior by grouping devices according to their attributes, which could include descriptive data and metadata like user age, model, and manufacture date.

For example, in the case of a fleet of connected elevators, a number of them may be experiencing too much downtime each month. Upon undergoing cluster analysis, data might point to a certain subset of the elevators that are problematic based on similar characteristics that include having been serviced more than twelve months ago and exceeding average runtimes of eight hours each day. Cluster analysis provides the elevator manufacturer with a reliable and efficient course of action: to schedule maintenance on elevators approaching twelve months since they were last serviced, prioritizing those running more than eight hours a day. Understanding this correlation also allows the elevator manufacturer to ensure that these problems do not arise in the future.

3.3.2 CLASSIFICATION

Classification is one of the most applicable concepts to IoT product providers when trying to predict device failure. Classification entails grouping device behavior by outcome into buckets, the number of which

varies based on the specific application.

To continue the elevator example, cluster analysis suggested that elevators maintained within a certain time frame, and that run over recommended thresholds each day, will have more downtime than is preferred. Imagine that an office building has four elevators and, when they are all running, none exceed eight hours a day of runtime. However, when one elevator is being repaired or replaced, one or more of the other elevators will have increased runtime, sometimes beyond the threshold of eight hours.

This would be cause for concern if an elevator was serviced more than twelve months ago, but otherwise is not a problem. A classification algorithm is capable of taking all of these variables (and many, many more) into consideration and can notify the appropriate party in the event that a combination of variables is met that indicate significant downtime could be imminent.

By dividing these outcomes into different behavioral buckets, unexpected downtime and loss of time, money, and customer satisfaction can be avoided.

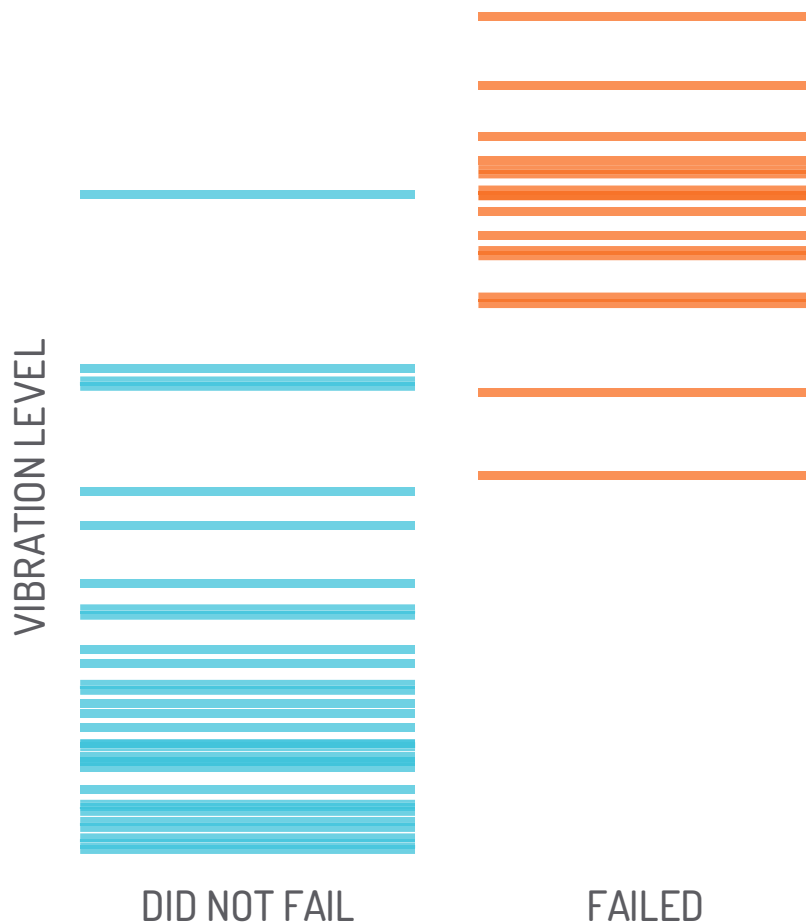


FIGURE 3.2: PUMP VIBRATION LEVELS AND FAILURE RATES

Based on this comparison, vibration level alone clearly does not predict device failure. Abnormally high vibration has several causes and, although most pump failures are caused by high vibration, not all high vibration is an indicator that a pump will fail.

In the case of predictive elevator maintenance, the stakes are likely increased wait time for elevators or additional downtime of elevators with higher daily runtimes, but in many cases the potential hazards of not knowing when a product will fail are less benign.

In industrial applications of predictive maintenance, it is common to simplify classifications into two buckets. To discuss this, consider a more detailed example based on an industrial fluid pump. For this product, the two classification buckets might be *pump will fail* and *pump will not fail*.

Through diagnostic analysis, it is possible to determine that excessive pump vibration as monitored through a sensor on the pump is the most common



cause of failure. Figure 3.2 provides average vibration data points from a pump, sorted by pumps that failed and those that did not fail within a period of time.

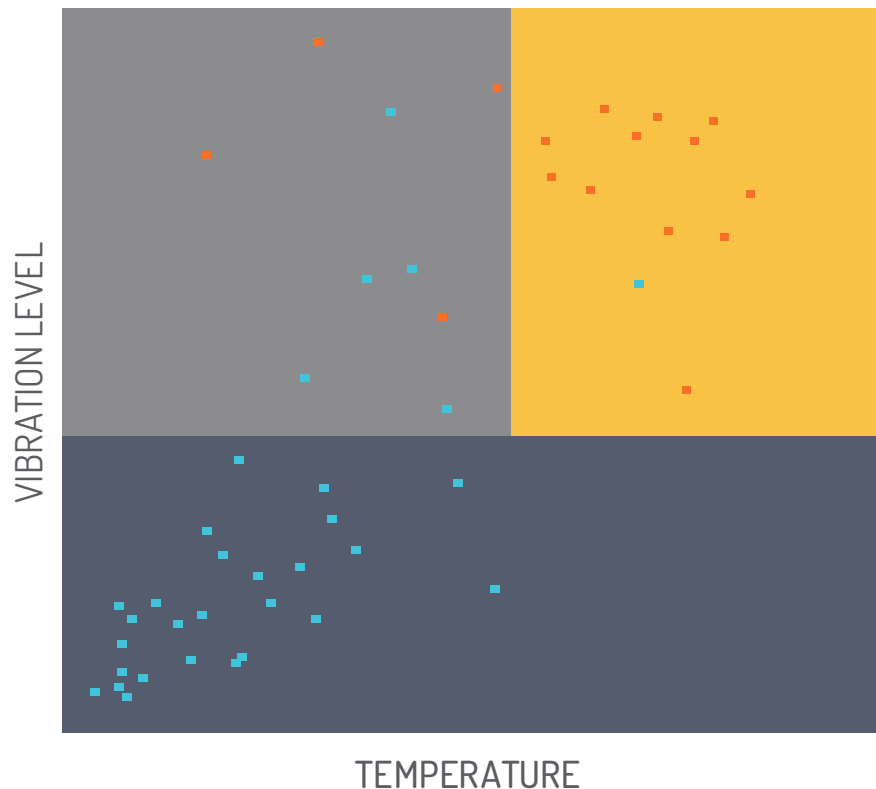
Based on this comparison, vibration level alone clearly does not predict device failure. Abnormally high vibration has several causes and, although most pump failures are caused by high vibration, not all high vibration is an indicator that a pump will fail.

This suggests that additional pump metrics should be considered. Figure 3.3 below compares average vibration amplitude and average temperature, which can also be monitored by a sensor on the pump, over a period of time on each axis.

FIGURE 3.3: PUMP VIBRATION VS. TEMPERATURE

When the temperature dimension is added, it is clear to see that high pump vibration level, coupled with high temperature, is an indicator of pump failure. However, it then becomes less obvious about how to explain the data points in the grey area through vibration and temperature alone.

 FAILED
 DID NOT FAIL



When the temperature dimension is added, it is clear to see that high pump vibration level coupled with high temperature is an indicator of pump failure. However, it then becomes less obvious about how to explain the data points in the grey area through vibration and temperature alone.

In order to do this, it may be useful to consider additional pump information, like the main cause of pump vibration. Say that excessive pump vibration is most commonly caused by bearing failure for a certain application. Further, if a bearing on a pump goes out, more friction, and thus heat, will be introduced to the system. If pump vibration is the only variable taken into consideration, a pump may be considered likely to fail because the vibration amplitude is above the typical range. However, if pump temperature is within its normal range, a bearing likely has not failed and the pump should not be

predicted to experience downtime. Increasing the sophistication of predictive models so that they take into account multiple variables and internal product knowledge can help avoid such false positives.

However, the more variables that are added to the mix, the more complex the logic behind predicting failure becomes. For instance, in pumps that have not been serviced in the past six months, high vibration indicates a problem unrelated to bearing failure even if temperature may be normal. For these logical, decision-making processes to be automated, it is necessary to build classification trees and encode them into predictive maintenance algorithms. Figure 3.4 below shows an example of a classification tree that takes into account the dimensions of vibration, time since last serviced, and temperature.

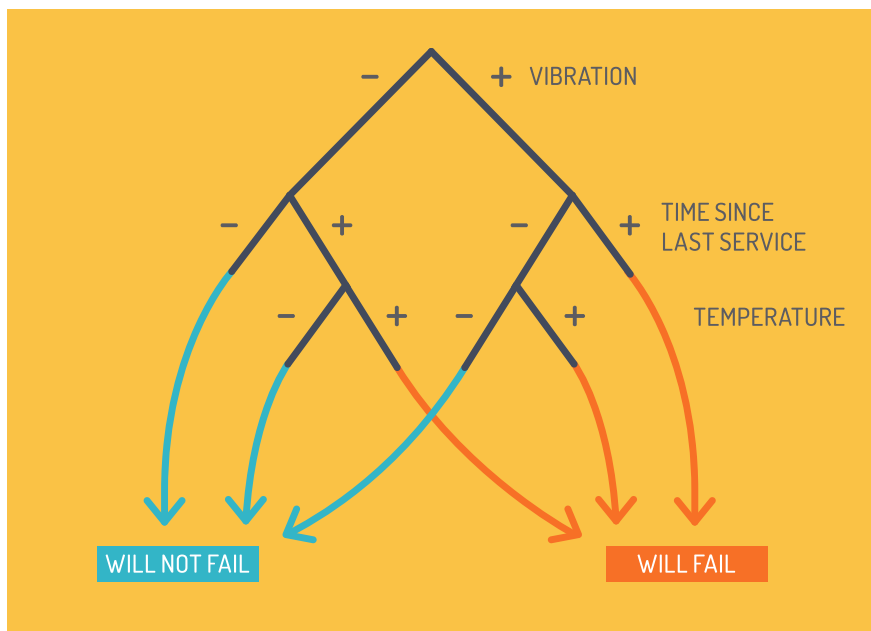


FIGURE 3.4:
PUMP
CLASSIFICATION
TREE

Using these three dimensions, the classification tree segments pump behavior into the buckets of *pump will fail* and *pump will not fail*.

Using these three dimensions, the classification tree segments pump behavior into the buckets of *pump will fail* and *pump will not fail*.

Typically, the simple model just created would be built by intensive statistical learning processes that create hundreds or thousands of models and find one with the best fit. However, the best fit for 100 pumps on one site might not be the best fit for 250 pumps at another. One of the primary dangers in using statistical techniques to construct a classification tree is overfitting. Overfitting occurs when a set of data is used to develop a classification of device behavior and the model is built to resemble that particular set of data as closely as possible. However, since runtime data will never be exactly the same as the sample used to create the model, overfitted classifications tend to not account for true relationships between variables and instead are built upon nuances of the test data set, many of which are random.

3.3.3 RANDOM FORESTS

Random forests are useful in combatting overfitting. A random forest involves the construction of many classification trees built on many subsets of a sample data set and averaging their results. This provides much more stable predictive power than an overfitted model does. As random forests frequently contain over 100 trees, it is outside the scope of this paper to discuss them in detail here.

3.4 PRESCRIPTIVE ANALYTICS: WHAT SHOULD BE DONE?

The final stage of analytics maturity involves deriving actionable items from predictions made in previous stages. When thousands of devices are online and their data is being routed through various machine-learning algorithms, they can provide users and administrators with incredibly valuable insights into what the fleet is doing and what action should be taken to modify its future behavior. The previous descriptive, diagnostic, and predictive analytics took the process one step closer to no-touch decision making—prescriptive analytics take it the rest of the way.

3.4.1 DECISION SUPPORT VS. DECISION AUTOMATION

Efficient and effective decision support and/or decision automation is the culmination of a mature data analysis program. Complex systems of information that would take months of analysis to understand can be streamlined into a decision-making algorithm that runs in a fraction of a second. However, the role prescriptive analytics play in supporting decisions or automating them depends heavily on the application.

Some applications call for results that support and inform decision makers so that they can efficiently make the best possible decisions for the company and device fleet. In the elevator example, decision support was provided in the form of a prioritized maintenance schedule for the elevator fleet, leaving the actual scheduling to a human who may have a more intimate understanding of different criteria for

which products need to be serviced first.

Other applications may include problems that have greater complexity or require faster response times and are, therefore, automated so that no human touch is required. A device that is about to fail and may cause injury, perhaps a pump on a load-bearing crane, will shut down safely and a replacement part will be ordered automatically.

4. TURNING INSIGHT INTO ACTION

For enterprises building connected products, collecting data is not enough. Data without context is meaningless. There is no one-size-fits-all data analysis solution and the role of descriptive, diagnostic, predictive, and prescriptive analytics will change for each application.

However, utilizing in-house expertise on products, services, and markets in collaboration with external analytics teams will enable an organization to develop the analytics program best suited to drive its business intelligence. In order to find business-changing insights in data, build a standardized data collection process by asking these three things:

What are your unanswered questions? Whether you aim to predict when a device will fail, understand how often a product should be serviced, or understand user segmentation, an analytics program can give you the answers you need to drive your business behavior and maximally monetize your IoT solution.

Do you have enough data to answer those questions? If not, what else do you need to measure? Understanding what data you need is an imperative first

step toward building out predictive models around your IoT solution. In collaboration with product experts, gathering a baseline set of metrics and data can yield process and product efficiencies that were previously unheard of.

How can you derive action from data? Deciding what kind of analytics strategy best suits your business case can help you eliminate, or all but eliminate, the level of human interaction required to make important decisions, from shutting off a failing piece of machinery to sending in a repair request.

Having a mature data analysis program is becoming table stakes for companies building connected products. Closing the gap from information to decision is one of the most important strategies for companies to utilize IoT to transform their business. As the number of connected devices increases into the tens of billions and costs decrease, companies taking charge of their data analysis and finding ways to monetize it will leap ahead of their competition and drive greater profits.

USA HEADQUARTERS

275 Market St, Suite 535
Minneapolis, MN 55405
+1.612.353.2161

TAIWAN OFFICE

WenXin Road, Section 4
#955, 15F-5
Taichung, 406 Taiwan
+886.4.2247.1623

REVOLUTIONIZE YOUR BUSINESS WITH IoT.

Engage with our team of world-renowned experts to learn more.

EXOSITE.COM | +1.612.353.2161