SIGHT MACHINE WHITE PAPER

Big Data B.S. Myths and Misinformation About Big Data for Manufacturers

Quick Take

- Separating Big Data “fact from fiction” is imperative for today’s manufacturer, given the often-misleading information that is now so abundant.
- This paper takes on two of the biggest myths in today’s market – involving both oversimplification and vendor claims.
- The paper also highlights a Big Data framework noting what Big Data is and isn’t, and what makes Big Data so challenging for the manufacturing enterprise.
- See how to distinguish the many claims, gain understanding as to what Big Data truly is, and how it impacts your business.

SightMachine.com | 888.461.5739 | info@SightMachine.com

White Paper

SIGHT MACHINE
Manufacturing Analytics

Big Data for the Manufacturing Enterprise

Leading manufacturers use Sight Machine's Big Data tools and technologies to capture, analyze, and consume data in new ways not seen before, making their processes more efficient and improving product quality. The solution is specifically designed for manufacturers, with data models and applications optimized for the real world of manufacturing.

As so often happens when new technologies emerge, some don't understand them, leading to a superficial and misleading conversation. That's been true of Big Data – it's been one false fact after another. It's time to bring the hammer down, because the true advantages of using data for manufacturing don't need to be embellished or exaggerated.

This paper is designed to address two of the biggest myths we see in the market.

Myth #1: Big Data is Magic – just build the lake...

The first myth is one of over-simplification. The story goes that manufacturers just need to build a giant data lake and then install Hadoop or another well-known processing framework, plus a business intelligence tool, and magic will start to happen.

Good analysis requires more than a lot of data. The scientific method involves observing the world, formulating hypotheses, gathering data, testing and interpreting results, publishing and replicating studies, etc. Each of these steps brings new and important information to the analysis.

> Big Data mythology has attempted to replace causation with correlation.

For example, creating a hypothesis brings prior knowledge and experience into the analysis and helps to reduce the risk of spurious correlations. Likewise, an understanding of the manufacturing processes helps to guide the best data analysis, rather than relying on algorithms that may over-fit the data. Rigorous analysis – grounded in theory, presented with strong visualizations – creates a confidence in the results, which is necessary when studying manufacturing processes that require costly changes and carry a high price of failure.

The Big Data mythology has attempted to replace causation with correlation. Granted, causation was certainly never easy to find in traditional analysis. But, carefully designed experiments efficiently and effectively explore the problem space, helping to identify root causes. In contrast, the data lake approach is often passive, limiting the range of inputs in the analysis, and thereby also limiting the range of insights that the analysis can produce. Fundamentally, data analysis is inductive reasoning, which means that the analyst is exploring for generalizable rules from specific data points. Intentionally designing experiments helps ensure a strong basis of starting data points from which to create powerful inferences.

Of course, that doesn't mean none of the Big Data methods have merit. The goal should be to triangulate between traditional analysis tools and the new opportunities created by Big Data. For example, Big Data anomaly detection can be paired with traditional Statistical Process Control tools to help surface interesting trends that merit further research. Algorithms can search potentially interesting correlations and present those results to the analyst, who can then refine the data and identify the models and relationships. Automated tools can help to categorize and organize the data and suggest appropriate analysis tools, helping non-data scientists perform rigorous analyses.

Also consider that Big Data has a strong history of translating analysis results into visually appealing graphics, helping to communicate the results to colleagues and management.

Myth #2: We're Big Data, Too

The second myth is one being propagated by legacy software vendors who suddenly claim to offer the benefits of Big Data. A company enables its users to see a report on the Web, and automatically considers itself a "cloud" company. This is NOT a Big Data company.

How can you tell the difference?

It's useful to begin with the definition of Big Data itself, popularly known as the three Vs: Volume, Velocity, and Variety. This framework highlights what it is exactly that makes Big Data different: a lot of data, processed quickly, of a wide variety. It's the combination of the three aspects that's new, and of the three, the most difficult problem is Variety – a problem that has been especially challenging in manufacturing. Legacy technologies address some of these challenges singly.
As an example, many software enterprise vendors have been able to compute large volumes of structured data for years – but dealing with just one of the 3Vs by itself isn't much of an advance. In fact, the structure of these older systems often bogs down in performance, runs only very specific queries and has no intelligence that enables them to improve over time.

> The wide variety of Big Data is the most difficult problem, especially challenging in manufacturing.

Here are some ways to figure out if a company is truly a Big Data solution:

- If companies are forced to limit the amount of data points they're pulling into an analysis for cost or performance reasons, they are NOT using a Big Data tool.

- If they're running entirely on an SQL server or another legacy relational database, they are NOT a Big Data company. They are not dealing with the Variety challenge – the core challenge for manufacturers.

- If the company is not prepared to understand and assure data quality, the company cannot be relied upon and is NOT a Big Data company. It's one thing to move volumes of data from your automation equipment into a database. It's altogether another to convert that data into useful insight that you can

rely on.

- If reports are run against a "report" database that is updated once every 12 or 24 hours, they are NOT a Big Data company.

- If users click "Run" and then walk away from their desks for a few hours (or overnight!) waiting for the results of a report, they are NOT using a Big Data solution.

Sight Machine: Big Data Analytics for Manufacturers

The analytics platform must solve more than a "manufacturing problem." Instead, we see first-hand evidence of a true Big Data problem, with the accompanying misunderstandings as outlined above.

Sight Machine's approach to harnessing the power of Big Data is marked by the following innovations:

Ubiquitous Data Collection: Tackling the wide variety of manufacturing data is a business imperative, as noted above. Every machine, every scanner, every sensor puts out a steady stream of data, both structured and unstructured. The analytics platform must capture and collect data from a wide variety of sources and formats, including: OPC PLCs, sensors and sensor-enabled robots, gage cameras, serial connections, Modbus, Windows shares, CC-link, Cognex, TCP/IP connections.

Cloud-Based: With advanced manufacturing analytics, these data are collected by a local data node (e.g., an industrial PC), and fed to cloud servers at appropriate times. The platform approach utilizes reusable components and a scalable data collection framework to enable a truly economical implementation. Web-based interface and cloud-based storage and analytics provide maximum flexibility.

Advanced Data Management: We believe in applying best manufacturing practices, such as standardized adapters for legacy MES, ERP and SCADA systems, along with advanced data management practices such as signal processing of sensors and images to higher level data structures. Additionally, it's key to access the latest advanced data tools such as supervised and unsupervised learning techniques to extract high-level insight from raw signals. Combining these approaches with a cloud-based data store that is always available enables real-time insight from both structured and unstructured data.

Ease of Access: Web and mobile-based applications must be available on demand, at any time, from virtually any connected device. The applications should be updated automatically, with comprehensive data collection and real- time results. Additionally, the application framework should enable historical reporting, predictive analysis, benchmarking, and data mining.

Data Output: In addition to providing analysis of the manufacturing data within the presentation layer, the platform should also integrate analyses and reports into ERP, MES, and other enterprise solutions, via a standard set of APIs (application programming interfaces).

In all, modern manufacturing analytics exist to improve understanding, and to enhance a company's

competitive position, using the Big Data flowing through and around the enterprise. Leading manufacturers use Sight Machine's Big Data platform to capture, analyze, and consume data in new ways not seen before, making their processes more efficient and improving product quality. The solution is specifically designed for manufacturers, with data models and applications optimized for the real world of manufacturing.

Today's manufacturing company needs powerful, practical technology to make sense of the flood of data previously locked behind complex and hard-to-use tools - and Sight Machine does just that.

SightMachine.com | 888.461.5739 | info@SightMachine.com

About Sight Machine

The Sight Machine Manufacturing Analytics Platform analyzes existing manufacturing data for trends and important statistics, presenting real-time information to manufacturers in an easily digestible cloud-based format. Integrated into existing manufacturing operations, Sight Machine applies best manufacturing practices, such as standardized adapters for legacy MES, ERP and SCADA systems, along with advanced data management such as signal processing of sensors and images to higher-level data structures.

Sight Machine's web and mobile-based applications are available on demand, at any time, from virtually any connected device – from across the plant or across the globe.