Data Science and Digital Manufacturing

Getting business value from factory floor data with Sight Machine

SIGHT MACHINE

The Data Science Challenge

Manufacturers are inundated with data. Thanks to the proliferation of cheap sensors, networked automation and control systems, and tools for acquiring raw data, there is no shortage of manufacturing data. The question is, now that you have it, what do you do with it?

To derive actionable information from raw manufacturing data, data scientists face four major hurdles:

1. **Managing a variety of high- and low-speed data streams that must be blended together.**
   A model of the manufacturing floor combines data including raw sensor streams from a robot sampled multiple times per second, ERP or MES data provided once per part, quality data collected per batch, maintenance data collected per shift, (hopefully) rare warranty claims data, etc. Then all these different speeds of data traffic must be merged in real time into a shared repository to build analytical models.

2. **Blending part and process data to create a complete picture of manufacturing operations.**
   Historically siloed data types should be combined together to provide the most information to feed into the analytics. For example, drifting measurements of parts in a QA lab are a predictor for preventive maintenance for a machine, or sensor data from an upstream device may predict downstream quality issues.

3. **Maintaining flexible data and statistical models that scale with the manufacturing process.**
   Although standardized mass scale is often a goal in manufacturing, actual production processes are constantly being improved, devices on the plant floor may be similar but not identical, or identical physical machines may be programmed to create different products. Data models and analytical tools must be able to flex with the manufacturing process.

4. **Building trust in the data and analytics through accurate, justifiable results.**
   Even the most sophisticated analyses are worthless if managers and operators can not trust the results and use them to drive improvement. Trusted analyses begin with good, clean, verified data streams. It is then enhanced by building upon more intuitively trustworthy manufacturing-centric data models such as machine cycles, downtimes, defects, parts, batches, etc., rather than relying on algorithms that combine high-variety data streams in black boxes that magically produce insight.

Resolving these challenges ultimately places a heavy burden of data pre-processing "grunt work" on data scientists. Instead of applying their specialized knowledge and skills to deriving manufacturing insights, they frequently spend a majority of their time on basic data processing, data blending, and data verification tasks.

Sight Machine's Approach

The data science process can be incredibly time consuming and resource intensive, and unfortunately, the application of the data is rarely multi-use. More often than not, requested modifications and new iterations require a repeat of the process, additional data acquisition, and additional coordination from stakeholders within the enterprise.

Sight Machine takes a different approach to manufacturing analytics. It is an approach streamlining data science activities by bridging the gap between raw data sources and the actionable analytics by building digital twins in the shortest amount of time with the least amount of pain. At Sight Machine, we've developed frameworks, heuristics, and patented automations to reduce the headache of data management, while ensuring all end-product analytics are cross-functional, reusable, and purpose-built for manufacturing.

STEP ONE: ACQUIRING AND PREPARING DATA

The obvious first step in any analytics process is acquiring good data to serve as the input into later statistical and machine learning analyses. Getting a lot of data is important. The more data you have available, the greater the chance that the solution can be found in the data. As the famous statistician John Tukey said, "The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data."

Unfortunately, acquiring data is also one of the most demanding processes in manufacturing data science contexts today. In a typical data acquisition process, the data team must:

→ Pull data from multiple raw data sources available across a plant or factory.

→ Within each source of data, there can be a wide variety of data streams, types, and formats, such as PLC, industrial control PCs, historians, databases, etc.

→ Historic data sets may not have been verified or tested, resulting in large volumes of worthless data.

→ Late arriving data, missing data, operator data entry issues, and related problems may create problems that are amplified when streaming and processing real time data.

→ Once data is accessed and formatted, extracts all too often can be used only for a single project or application.

The Current State of Data Acquisition

The data acquisition process typical of many manufacturing enterprises and their data science resources is characterized by three steps, each with its own challenges:

→ **Accessing Data Streams:** Currently, data acquisition requires that data scientists tap into a variety of raw data streams. In some cases, this is a relatively straightforward practice. For example, all that may be required is accessing data in MES and ERP systems that sit on top of other databases. In other cases, this is a more complicated exercise, requiring substantial heavy lifting to even access data, such as directly acquiring data from a PLC or working with operator entry data fields containing free text.

→ **Verifying Sources, Types and Formats:** Once these streams are tapped, the data forensics process begins with data scientists parsing and verifying manual sources to structure and format consistent data. This is where the grunt work of data science builds its reputation for intensive data munging and modeling, often for a single-use application.

→ **Keeping Data Updated:** Once data is prepared and ready, keeping data updated becomes the primary challenge – without it, repeat analysis and future improvement is severely limited. Throughout a manufacturing enterprise, data flows at different intervals based on the source. Maintaining updated data or establishing a continuous extraction proves too difficult, and this process is repeated case-by-case for a given model, application, or analysis.

Sight Machine's technology condenses these steps into a 3-stage pipeline building process that ensures that data flows can be used repeatedly. As an end result, data is conditioned for ongoing digital twin use, and the grunt work of one-off data conditioning is substantially reduced. The established data pipeline enables data scientists to focus more time on high value-add analytics instead of basic data preparation.

SIGHT MACHINE'S 3-STAGE PIPELINE PROCESS

Sight Machine's approach for conditioning data to be used for digital twin building consistently follows a 3-stage pipeline process. Refined from years of work with enterprise manufacturers, this approach elevates the importance of time stamped machine data with appended metadata that can be consistently interpreted for future, repeat applications.

STEP 1
Extract data from original sources and pull it into the local system. These sources can widely vary to include queries from SQL, files from FTP server, files from window shares, images from machine vision systems, tags from historians, and registers from a PLC.

STEP 2
Combine a standard data payload of the timestamp and metadata to identify characteristics such as location within a plant, any connected assets, etc. Then, begin to classify data in terms of "type." An example could be a time series that shows a complete picture at every step in time, a list of events, and end-of-batch reporting, describing an entire shift's worth of activity in aggregate format.

STEP 3
Record the combined payload of original machine and meta data projected upon it into a consistent format using JSON. Then, push the data payload upstream to either a gateway or to the cloud.

Sight Machine reduces data acquisition headaches and expands the analysis capability beyond the boundaries of automation integrations. As a result, data scientists have access to more power with data in the cloud and can allocate more resources to run the analytics.

STEP TWO: CONTEXTUALIZING DATA

Sight Machine then contextualizes the raw data stream into manufacturing data models to build a digital twin. This helps to build trust in the data by converting it into familiar forms, and shortens the gaps between raw data streams and actionable manufacturing insight. Once data is acquired, prepared, and ready for use, our patented AI Data Pipeline establishes a baseline of consistent, inbound data. This solves the problem that many manufacturers encounter due to segmented workflows and provides a new level of data connectivity across the entire operation.

From years of familiarity working with traditional manufacturing analytics such as Six Sigma and Lean production, Sight Machine has developed heuristics and algorithms specific to manufacturing. This informs the digital twin technology that applies context to data in a way that is virtually impossible to manually build and scale across an enterprise.

By establishing the data stream for ongoing data contextualization and digital twin building, Sight Machine allows for inbound data to easily translate to manufacturing and business audiences. With Sight Machine's technology providing a holistic view of performance, trust is built cross-organizationally about the value of data, while the pressure is off of data science departments to produce single-application, contextual analyses, and models for stakeholders.

Sight Machine's Out-of-the-Box Digital Twin Technology

Most data science tools are not end-to-end. By streamlining the workflow and limiting time spent on producing data, Sight Machine places the emphasis on analysis. This allows data science teams to auto-run analyses and focus on understanding the solution from the business perspective.

Many of our out-of-the-box implementations follow a familiar sequence of steps to a proof of concept before scaling to larger factory and system-wide digital twin applications:

→ Work with manufacturers to understand a specific production problem.

→ Form a series of hypotheses about possible causes with manufacturers.

→ Using **Sight Machine's Analytics Workbench**, define and input data into the platform to support the analysis.

→ **Sight Machine's Root Cause Analysis Tools** run a series of heuristics to narrow down the problem space and reduce the number of variables.

→ Predictors are narrowed down through built-in algorithms designed to select the best statistical approach based on input/output data.

→ Likely variables of interest and final predictors are presented to the analyst for further evaluation and potential process change.

Sight Machine's out-of-the-box technology meets manufacturers' needs for a general-purpose, reusable software platform while also being able to solve specific production challenges. The end result is a series of abstract concepts that Sight Machine makes concrete. Data scientists and statisticians can then combine their domain expertise with the likely variables of interest or predictors to follow up on potential causes to the problem.

STEP THREE: ANALYZING & VISUALIZING DATA

With Sight Machine's rolling stream of collected data from defined sources, manufacturers elevate the base level of performance and enable shared tacit knowledge to be visible and distributed throughout the entire enterprise. Purpose-built for the industry, Sight Machine understands the data challenges facing data scientists within manufacturing enterprises. Harnessing the power of Big Data within the manufacturing context, Sight Machine converts raw data inputs into manufacturing insights.

By enabling data scientists to easily replicate and scale existing digital twin models, Sight Machine's technology stands apart from others by providing industry-specific value to manufacturers. Equipped with plant, factory, line, machine, part, and batch combined insights to solve production and operational problems, data scientists have proven tools for delivering value to business stakeholders.

Data Science Trends & Directions

THE FUTURE OF DATA SCIENCE

The data science landscape within the manufacturing industry continues to evolve. The rise of automation equipment and machine telemetry in the last decades introduced the possibility of big data for manufacturers. Now housing nearly 20% of all data across all industries, manufacturing as an industry is a data powerhouse. Sourcing the data science resources to maintain and make sense of these data streams is an increasing demand.

Currently, many manufacturers are working to define the scale of their internal data science resources, often pulling from production and manufacturing engineer talent pools within their organization. Others have built extensive data science departments to support data-oriented decision making. Regardless of the scale of manufacturers' existing data science capabilities, the complexity of statistical models is advancing and will continue with the increase of data science resources supporting manufacturing analytics.

As the only purpose-built manufacturing analytics technology on the market, Sight Machine's technology for building digital twins provides unmatched value to enterprises globally. More and more data experts are entering the field to move today's advanced statistical models toward increasingly physical models. Sight Machine's technology, built to establish a powerful baseline of contextual data for manufacturers, is positioned to support these more complex models of the machine.

CONCLUSION

The multi-step process of converting a raw data inputs into valuable, manufacturing insights for a specific end user and application is one that data scientists are familiar with executing. Many software developers and technology providers have identified opportunities to support pieces and parts of this process. Sight Machine takes a different approach.

With an extensive knowledge of manufacturing, Sight Machine's technology is purpose-built to support each step of the data collection and contextualization process. While many other products will do the data wrangling and blending or the dashboard and report building, Sight Machine creates manufacturing models as inputs for analysis that data scientists can use to create manufacturing insights as the output for business stakeholders to drive decisions. This is an end-to-end solution that data scientists can use to provide manufacturing insights that improve performance.

To learn more about Sight Machine's digital manufacturing technology, contact us at
info@sightmachine.com