



Data-Driven Manufacturing: AI, ML, and Data Science

Dr. James A. Bednar
Director, Custom Services

Manufacturers are awash in data!

- Sensors are cheap – you probably have loads of them already!
- Plus you have lots of data on your inventory, supply chains, and markets
- Making intelligent decisions based on your data can give you a major leg up
- McKinsey estimates that \$1-2 trillion in manufacturing revenue is on the table



“Manufacturing” covers a huge range of activities

- **High-volume, high-tolerance mass production of commodities:**
Need data to reduce material and energy costs, increase yields, predict maintenance
- **Low-volume, low-tolerance precision manufacturing:**
Need data to detect defects, improve performance, optimize processes
- **Mass customization:**
Need data to drive computer-controlled machinery and processes
- **Fully custom shop:**
Need data to meet specs and determine price points, feasibility

No single approach is going to meet all these needs!



So... how do you extract value from your data?

If you're Facebook/Amazon/Apple/Netflix/Google/Microsoft, you

- hire world-leading data-science and machine-learning/artificial-intelligence experts
- push the boundary of what is possible for computers and software to achieve
- build on nation-state-level data infrastructure

Is that the right approach for you, though?



Don't hire a big, expensive AI/ML/DS team with grand goals

- UNLESS you already have extensive data infrastructure in place
- AND have an extensive team of subject-matter experts (SMEs)
- AND the SMEs are motivated and officially charged with making this project a success
- AND you have the power to rebuild the organization's decision-making processes to be driven by data

Without all those in place, success at ambitious goals is **highly unlikely**.



Don't just throw ML at your problem!

- You can find all sorts of online tutorials and “auto ML” tools that promise instant insights from arbitrary datasets
- Across many customers, industries, and projects over many years, we are not aware of that promise ever having been realized, even with newer LLM/generative-AI models.
- Main problem: not understanding the data, the underlying system, and how they relate
- Insight from a raw dump of numbers is very unlikely; what's signal and what's noise?
- Conversely, domain experts don't typically have the data-handling, interpretation, and coding skills needed to work with large and complex datasets



Instead, make a small, feasible plan

- **What is your goal? Be very specific about your desired outcome.**
(Hint: Pick something feasible to start with, e.g. improving one particular process)
- **Is this a one-time need, or ongoing?**
(Hint: Start with a one-time need until you get up to speed; data pipelines are hard)
- **Do you already have the data you need?**
(Hint: Start over with a different problem, if you don't yet have data for the one you want)
- **Do you have the organizational will to make changes?**
(Hint: Use a small project to demonstrate results and get buy-in, first!)
- **Do you have people who understand the process/system AND how to work with data?**
(Hint: It won't work if you don't, but they don't have to be the same people.)



How Anaconda helps manufacturers

- Anaconda as a company is built around open source (OSS) Python libraries:
 - Packaging OSS tools
 - Helping you building OSS-based environments for computations
 - Expanding and advancing OSS tools for working with data
- We help companies establish robust OSS-based computational pipelines from raw data to visualization/analytics tools to automated processes
- The first step in any of our projects (and often the only) step:
Let's take a look at (all) that data!



Example Anaconda/manufacturing projects (1/4)

Consumer beverage company analyzing production samples

Issue: “poor performance” from ML model of chemical makeup of sample from measurements

Solution:

- Visualized a set of black-box algorithms with HoloViews
- Found data being corrupted in compiled code early in the pipeline
- Simple non-ML algorithm sufficed after fixing that



Example Anaconda/manufacturing projects (2/4)

Oil and gas company making production decisions for wells from sensor data

Issue: data pipeline too slow to handle data volumes

Solution:

- Numba to compile Python to fast machine code
- Implemented efficient chunked data handling for incoming streams of measurements
- Panel app displaying data for decision makers



Example Anaconda/manufacturing projects (3/4)

Consumer device manufacturer optimizing for device performance

Issue: too much benchmarking data for standard visualization tools

Solution:

- Server-side rendering using Datashader to handle large data volumes
- Extensions to Bokeh and HoloViews to display many sources of information together
- Resulting analyses packaged as Panel dashboards



Example Anaconda/manufacturing projects (4/4)

Construction materials manufacturer looking to improve processes

Issue: Disconnect between SMEs with problems to solve and data scientists with skills

Solution:

- Set up OSS tools for working with data across the spectrum from:
 - non-coder SMEs (deployed dashboards and reports) to
 - analysts (low-code dashboards and preconfigured notebooks) to
 - programmers/data-scientists (Python)
- All tools connect directly to the same live data feeds
- All tools can share the same viz/analysis components whether full/low/no code
- Tools collect annotations from end users and display for humans or record for ML



Tools we use in our client work

OSS:

- **HoloViz.org**: Visualize *all* your data (always the first step!)
- **Panel.HoloViz.org**: Use Python to build data apps for use throughout your organization
- **Github.com/Panstacks/Pandata**: Scalable data-processing pipelines
- **MLFlow**: Keep track of all your experiments and model runs

Commercial:

- **Anaconda Distribution**: Secure builds of all the Python OSS packages you need
- **Anaconda Data Science Platform**: Kubernetes-based AI/ML/DS development and deployment, for building fully OSS data-processing pipelines (no vendor lock-in!)



The Pandata Stack: Scalable Open-Source Analysis

Data
storage



Data
access



Data
API

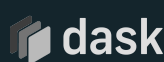


Awkward
Array



GRAPHBLAS

Data
processing

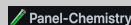


Your domain-
specific code



Dask-ML

dask-image



Visualization



User
interface



Packaging



Ok, I started small and it worked! Now what?

- **Great! Do you have the organizational will to expand?**
(If not, stay small while you work on the politics, consolidating your power)
- **Do you have the data you need to expand?**
(If not, is getting usable data feasible? It will always be more expensive and take longer than you think!)
- **Are there lots of other opportunities on specific processes?**
(If so, see if you can set up tooling that can be used across the board)
- **Or are you ready to start optimizing the whole factory's operations?**
(You'll need even more buy-in across the organization!)
- **Or are you ready to look outwards towards the whole supply chain?**
(Understanding inefficiencies, gaps, and comparative advantage could be a big win!)



Thank you and Q&A

- Thanks to all our amazing manufacturing clients over the years!
- Special thanks to our partners at the National Manufacturing Institute of Scotland, especially Prof. Andrew Sherlock, for helpful discussions!
- See the [Anaconda Blog](#) for more info
- Contact jbednar@anaconda.com or sales@anaconda.com if you want our help applying OSS to your own data problems!

