How To Build Great Data Products

Column	@August 9, 2021 9:14 PM
Due Data	@August 10, 2021
Readings	
Status	Done
Topic	🏆 Product Management

How to Build Great Data Products

Products fueled by data and machine learning can be a powerful way to solve users' needs. They can also create a "data moat" that can help stave off the competition. Classic examples

https://hbr.org/2018/10/how-to-build-great-data-products

Summary: *Products fueled by data and machine learning can be a powerful way to solve users' needs and stave off competition. Classic examples include Google search and Amazon product recommendations, but the opportunity extends far beyond the tech giants. To tackle the challenge, companies should emphasize cross-functional collaboration, evaluate and prioritize.*

The lifecycle of a so-called "data product" mirrors standard product development: identifying the opportunity to solve a core user need, building an initial version, and then evaluating its impact and iterating.

But the data component adds an extra layer of complexity. To tackle the challenge, companies should emphasize cross-functional collaboration, evaluate and prioritize data product opportunities with an eye to the long-term, and start simple.

Stage 1: Identify the opportunity

Data products are a team sport

Identifying the best data-product opportunities demands marrying the product-and-business perspective with the tech-and-data perspective.

Product managers, user researchers, and business leaders have the intuition and domain expertise to identify key unsolved user and business needs

data scientists and engineers identify feasible data-powered solutions and have the intuition on what can be scaled and how.

To get the right data product opportunities identified and prioritized bring these two sides of the table together.

The higher the data literacy of the product and business functions, the better able they'll be to collaborate with the data science and tech teams.

Prioritize with an eye to the future

The best data products get better with age, like a fine wine. This is true for two reasons:

1. Data product applications generally accelerate data collection which in turn improves the application. With limited profile data today, the initial (or "cold start") recommendations may be uninspiring. But if users are more willing to fill in a profile when it's used to personalize their experience, launching recommendations will accelerate profile collection, improving the recommendations over time.
2. Many data products can be built out to power multiple applications. It isn't just about spreading costly R&D across different use-cases; it's about building network effects through shared data. If the data produced by each application feeds back to the underlying data foundations, this improves the applications

Too much focus on near-term performance can yield underinvestment in promising medium- or long-term opportunities.

the criticality of high-quality data cannot be overstated; investments in collecting and storing data should be prioritized at every stage.

Stage 2: Build the product

De-risk by staging execution

Data products generally require validation both of whether the algorithm works, and of whether users like it.

As a result, builders of data products face an inherent tension between how much to invest in the R&D upfront and how quickly to get the application out to validate that it solves a core need.

Teams that over-invest in technical validation before validating product-market fit risk wasted R&D efforts pointed at the wrong problem or solution. Conversely, underpowered prototype, and so risk a false negative.

While there's no silver bullet for simultaneously validating the tech and the product-market fit, staged execution can help.

Starting simple will accelerate both testing and the collection of valuable data.

A series of MVP approaches can also reduce time to testing:

- *Lightweight models* are generally faster to ship and have the added benefit of being easier to explain, debug, and build upon over time. While deep learning can be powerful (and certainly is trending) in most cases it's not the place to start.
- *External data sources*, whether open source or buy/partner solutions, can accelerate development
- *Narrowing the domain* can reduce the scope of the algorithmic challenge to start. For example, some applications can initially be built and launched only for a subset of users or use-cases.
- *Hand-curation* — where humans either do the work you eventually hope the model will do, or at least review and tweak the initial model's output — can further accelerate development. This is ideally done with an eye to how the hand-curation steps could be automated over time to scale up the product.

Stage 3: Evaluate and iterate

Consider future potential when evaluating data product performance.

the data product may improve substantially as you collect more data, and because foundational data products may enable much more functionality over time.

Before canning a data product that does not look like an obvious win, ask your data scientists to quantify answers to a few important questions. For example:

- at what rate is the product improving organically from data collection?
- How much low-hanging fruit is there for algorithmic improvements?

Speed of iteration matters.

Data products often need iteration on both the algorithms and the UI. The challenges is to determine where the highest-value iterations will come from, based on data and user feedback, so teams know which functions are on the hook for driving improvements.

By fostering collaboration between product and business leaders and data scientists, prioritizing investments with an eye to the future, and starting simple, companies of all shapes and sizes can accelerate their development of powerful data products that solve core user needs, fuel the business, and create lasting competitive advantage.