

How to get started with machine learning

Asay, Matt . InfoWorld.com ; San Mateo (Aug 4, 2017).

[ProQuest document link](#)

ABSTRACT (ENGLISH)

Machine learning pipelines demystified [bullet] Google's machine-learning cloud pipeline explained [bullet] R and Python drive SQL Server 2017 into machine learning. | Keep up with hot topics in programming with InfoWorld's App Dev Report newsletter.] "Unless you already know about machine learning and how to bring it to production, you probably don't understand the complexities that you are about to add to your companies life cycle. The risks? "Boundary erosion, entanglement, hidden feedback loops, undeclared consumers, data dependencies, configuration issues, changes in the external world, and a variety of system-level antipatterns." Get started with Azure Machine Learning. | The InfoWorld review roundup: AWS, Microsoft, Databricks, Google, HPE, and IBM machine learning in the cloud.] But TensorFlow's (as well as Caffe's, Mxnet's, and CNTK's) shaking of the foundations of the machine learning orthodoxy is not the big deal, in Dunning's opinion. Real machine learning success, in other words, isn't going to come from an off-the-shelf software package, no matter how hard the company markets it as such (think IBM's Watson). [...]immediately dump the myth of the model.

FULL TEXT

Machine learning is still a pipe dream for most organizations, with Gartner estimating that fewer than 15 percent of enterprises successfully get machine learning into production. Even so, companies need to start experimenting now with machine learning so that they can build it into their DNA.

Easy? Not even close, says Ted Dunning, chief application architect at MapR, but "anybody who thinks that they can just buy magic bullets off the shelf has no business" buying machine learning technology in the first place.

[Learn how to write apps that take full advantage of machine learning: Data in, intelligence out: Machine learning pipelines demystified [bullet] Google's machine-learning cloud pipeline explained [bullet] R and Python drive SQL Server 2017 into machine learning. | Keep up with hot topics in programming with InfoWorld's App Dev Report newsletter.]

"Unless you already know about machine learning and how to bring it to production, you probably don't understand the complexities that you are about to add to your companies life cycle. On the other hand, if you have done this before, well-done machine learning can definitely be a really surprisingly large differentiator," Dunning says.

Open source projects like TensorFlow can dramatically improve an enterprise's chances of machine learning success. TensorFlow "has made it possible for people without advanced mathematical training to build complex -- and sometimes useful -- models." That's a big deal, and points to TensorFlow, or other similar projects, as the best on-ramp to machine learning for most organizations.

Machine learning for nothing, predictions for free

Machine learning success rates are so low because "machine learning presents new obstacles that are not handled well by standard software engineering practices," Dunning says. A successful dataops team involves complicated lines of communication and a multipronged development process.

Couple those complexities with the reality that machine learning systems "can easily have hidden and very subtle dependencies," and you have a perfect form for things going awry.

Google, which knows the payoffs and pitfalls of machine learning more than most, has written about the hidden technical debt imposed by systems that use machine learning. As the Google authors stress, "It is common to incur massive ongoing maintenance costs in real-world machine learning systems." The risks? "Boundary erosion,

entanglement, hidden feedback loops, undeclared consumers, data dependencies, configuration issues, changes in the external world, and a variety of system-level antipatterns.”

And that’s just for starters.

Not surprisingly, software engineering teams are generally not well-equipped to handle these complexities and so can fail pretty seriously. “A good, solid, and comprehensive platform that lets you scale effortlessly is a critical component” to overcoming some of this complexity, Dunning says. “You need to focus maniacally on establishing value for your customers and you can’t do that if you don’t get a platform that has all the capabilities you need and that will allow you to focus on the data in your life and how that is going to lead to customer-perceived value.”

Enter TensorFlow.

The four ways that TensorFlow makes machine learning possible

Open source, a common currency for developers, has taken on a more important role in big data. Even so, Dunning asserts that “open source projects have never really been on the leading edge of production machine learning until quite recently.” With Google’s introduction of TensorFlow, a tectonic shift began.

[Jump into Microsoft’s drag-and-drop machine learning studio: Get started with Azure Machine Learning. | The InfoWorld review roundup: AWS, Microsoft, Databricks, Google, HPE, and IBM machine learning in the cloud.]

But TensorFlow’s (as well as Caffe’s, Mxnet’s, and CNTK’s) shaking of the foundations of the machine learning orthodoxy is not the big deal, in Dunning’s opinion. No, “the really big deal is that there is now a framework that is 1) powerful enough to do very significant projects, 2) widely accepted and widely used, and 3) provides enough abstraction from the [underlying] advanced mathematics.”

His first point - the power to do real machine learning projects – is a gimme. Being limited to very simple models is not the way to stage an machine learning revolution.

His second point, however, is more surprising: “The point is that we need a system to be used by a wide variety of teams working on a wider variety of problems to have enough people write simple examples for newbies. We need a system that becomes a standard for accompanying implementations with machine learning papers so that we can tell where the paper glossed over some details.”

His third point about abstraction is also very important: “The fact that program transformation can produce code that implements a derivative of a function efficiently was not at all apparent even just a short while ago.” But it’s critical. “That capability, more than anything else – including deep learning – has made it possible for people without advanced mathematical training to build complex – and sometimes useful – models.”

With TensorFlow and other open source projects like it, teams can acquire new skills to successfully deploy machine learning by iterating and experimenting. This willingness to get hands dirty with open source code is his fourth point, that “successfully deploying machine learning will require that a team is willing to look deeply into how things work.”

Real machine learning success, in other words, isn’t going to come from an off-the-shelf software package, no matter how hard the company markets it as such (think IBM’s Watson).

Recommendations for doing real machine learning

For those that are ready to embark on a real machine learning journey, TensorFlow is a great way to get started. As you embark on that journey, Dunning has two recommendations:

First, prioritize logistical issues, a model delivery framework, metrics, and model evaluation. “If all you have is a model and no good data and model execution pipeline, you are bound to fail.”

Second, “immediately dump the myth of *the* model. You won’t have a single model for one function when it all gets into production. You are going to have multiple models for multiple functions. You are going to have subtle interactions. You are going to have to be able to run a model for quite some time to make sure it is ready for prime time. You are going to need to have perfect histories of input records. You are going to need to know what all the potential models would respond to input.”

Credit: By Matt Asay

DETAILS

Subject:	Success; Machine learning; Teams; Open source software; Artificial intelligence; Software engineering
Company / organization:	Name: InfoWorld; NAICS: 511120; Name: Google Inc; NAICS: 519130; Name: Microsoft Corp; NAICS: 334614, 511210
Publication title:	InfoWorld.com; San Mateo
Publication year:	2017
Publication date:	Aug 4, 2017
Publisher:	Infoworld Media Group
Place of publication:	San Mateo
Country of publication:	United States, San Mateo
Publication subject:	Computers--Microcomputers, Computers--Computer Industry
Source type:	Trade Journals
Language of publication:	English
Document type:	News
ProQuest document ID:	1925988207
Document URL:	https://csuglobal.idm.oclc.org/login?url=https://search.proquest.com/docview/1925988207?accountid=38569
Copyright:	Copyright Infoworld Media Group Aug 4, 2017
Last updated:	2017-08-05
Database:	ABI/INFORM Collection

LINKS

[Linking Service](#)

Database copyright © 2019 ProQuest LLC. All rights reserved.

[Terms and Conditions](#) [Contact ProQuest](#)