

So you want to build a ML system

Jan Macháek

May 27, 2018

Abstract

It concludes with a ML readiness test, and recommendations on how to roll-out machine learning to an organisation (including tips on the process to follow, the people to hire). Mature ML organisation should enable a new person to make an impact in a ML system on the first day!

1 I want ML

... TODO: ML Joel test

1. Versioned, testable; continuously tested and sanity-checked analytics (BI)
2. Any BI query can be answered under 10 minutes
3. Monitoring on the BI environment to identify queries that use normalised data
4. The results of the BI queries that humans process define what ML should solve
5. Ingestion components decoupled from the rest of the system
6. Versioned, testable; continuously tested and sanity-checked data sets
7. Pre-computed “return constant” model
8. Versioned, testable; continuously tested and sanity-checked model storage with training and validation data set references
9. Model deployer and “debugger”

But where is the ML that builds the model? That’s the code that the engineering teams need to build to replace item 6.

Once the first four steps are known, the engineering teams can implement the remaining steps of the pipeline. If the system that is to take advantage of ML is event-based, the event delivery mechanism provides the decoupling, resulting in architecture shown in Figure 1.

The ML team maintains the tooling for the pipeline, consults on the best models, researches, ...; but the product teams (that ultimately work on the service that *uses* the model) have the first dibs on implementing the model. Successful implementation of this strategy means that anyone can implement a new model (even if only to just see what will happen!), train it, debug it, and deploy it all within a single day. All the mechanics of data ingestion, storage,

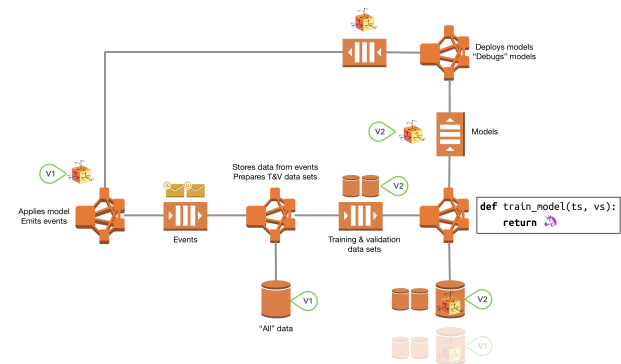


Figure 1: ML pipeline with code

versioning; runtime of training a model, evaluation, storage, versioning; debugging and deploying; and the usage is all implemented.

In this sense, the machine learning code is just like any other ordinary code; it is subject to all the high engineering standards and safeguards.