Building a Machine Learning Platform at Quora

The Quora Answer To “Build vs Buy” For ML Platforms

Nikhil Garg

@nikhilgarg28

@Quora

@MLconf 11/11/16

A bit about me...

- At Quora since 2012
- Currently leading two ML engineering teams:
  - Content Quality
  - ML Platform

@nikhilgarg28

Quora

The best answer to any question

To Grow And Share World’s Knowledge

What good things does the TPP hold for underemployed American citizens?

Will this trade agreement raise the U.S. workforce participation rate?

Answer | Request | Follow 6 | Comment | Share | Downvote | •••

9 Answers

Barack Obama, President of the United States

Written Thu · Featured in Inc · Upvoted by Marc Bodnick, Harvard Gov major, Stanford PoliSci PhD student, Nadia Singer, and 27 others you follow

The underemployed – folks who work part-time and want to work full-time – have a lot to gain from the Trans-Pacific Partnership, or TPP. I'm not the only one who believes this – a number of independent, credible analyses of TPP have shown it will grow exports, grow our economy, and raise incomes.

Underemployment and unemployment are driven by an economy that's not producing at its full potential – factories that aren't humming at capacity and workers who aren't being put to their highest and best use. It makes sense – fewer goods, services, and exports

What are the best Caltrain hacks?

Answer | Request | Follow 72 | Comment 1 | Share | Downvote | •••

5 Answers

Matt Laroche, Ex-Caltrain Regular.

Updated Feb 25 · Upvoted by Jeremy Lipps, I am the Social Media Officer for Caltrain., Nadia Singer, and 23 others you follow

Tickets/Passes

- Load your Caltrain monthly (and other passes/credit) at Walgreen's (or another Clipper retail location: https://clippercard.com/ClipperW...). It avoids the 3 day wait when you buy a pass online. This also means you can use your monthly without doing the zone based tagging in and out. (There's a few edge cases that fail with online purchase of monthly passes)
- Monthly passes are valid for all zones on weekends and holidays!
- 2+ zone monthly passes are treated, basically, as monthly passes for VTA and

Over 100 million monthly uniques

Millions of questions & answers

In hundreds of thousands of topics

Supported by 80 engineers

What Slows Down ML Innovation?

Curse Of Complexity

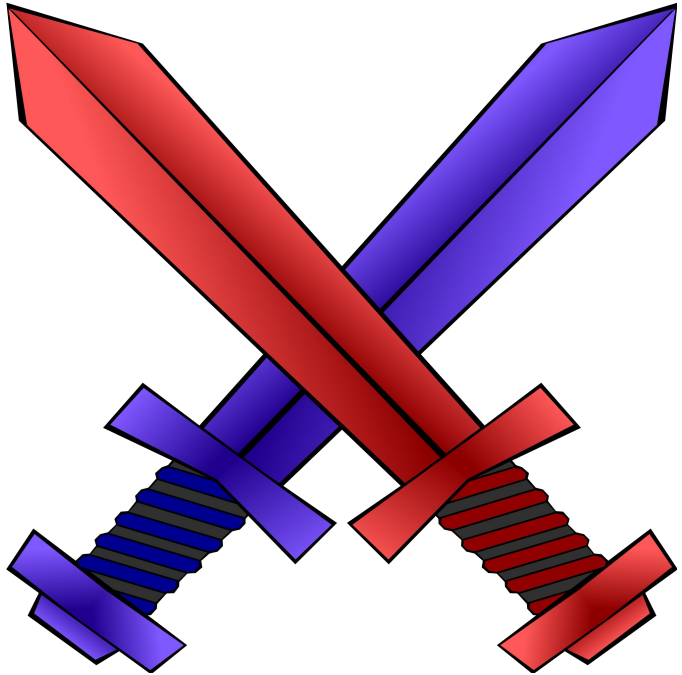
Quora

- Pipeline jungles
- Lots of glue code to get data in/out of general purpose packages.
- Strong coupling between business logic, data, ML algorithms and configuration.

Clash Of Titans

Quora

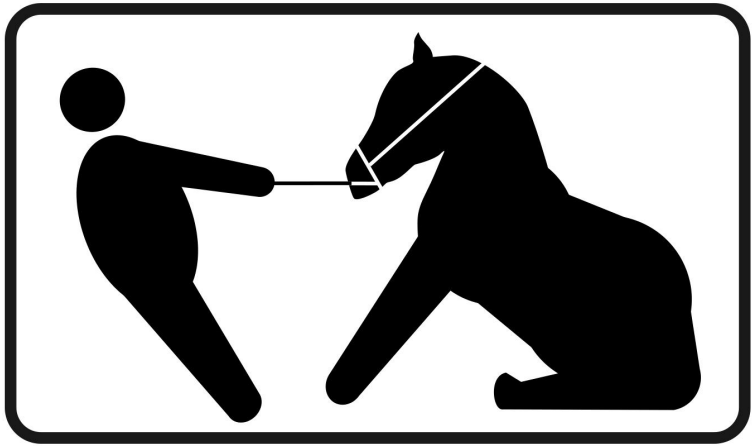
- Online vs offline
- Production vs experimentation
- C++ vs Python
- Engineering vs research
- ...even more glue code and pipeline jungles.



Getting New Applications Off The Ground

Quora

- Hard to reuse existing features, data, algorithms, tooling etc.
- Too costly to even get off the ground.



http://www.qvidian.com/blog/resistance-to-change-sales-organizations

Many Faces Of Chaos

Quora

Machine Learning:
The High-Interest Credit Card of Technical Debt

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov,
Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young
{dsculley,gholt,dgg,edavydov}@google.com
{toddphillips,ebner,vchaudhary,mwyoung}@google.com
Google, Inc

One ring to bring them all and in the darkness bind them!

Machine Learning Platform

Collection of systems to sustainably increase the business impact of ML at scale.

ML Platform: Build or Buy?

The Quora Answer: Build

For Seven Reasons

Reason # 7

Just Can’t Buy Everything!

Degree Of Integration & Delegation

Quora

- No matter how powerful the platform is, still need to maintain some form of integration
- This thin integration layer then becomes the platform.
- Real questions --
  - How much does this in-house layer delegate?
  - How much control does it have over delegation?

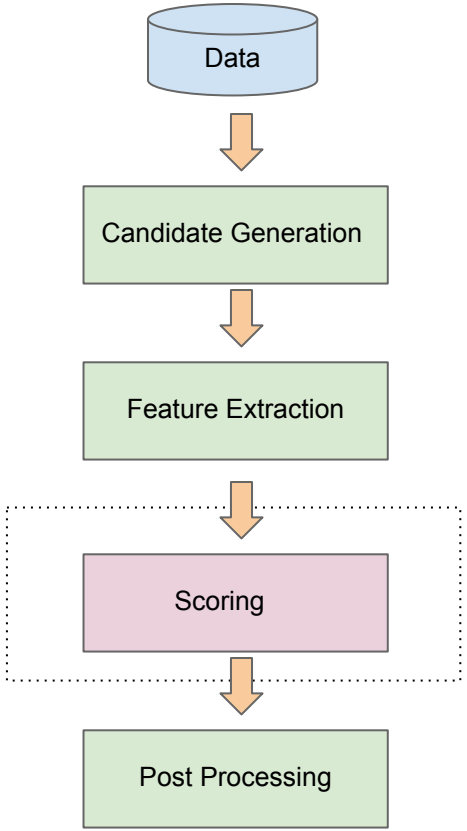
Reason # 6

Fast Scalable Production Systems

End-To-End Online Production Systems

Quora

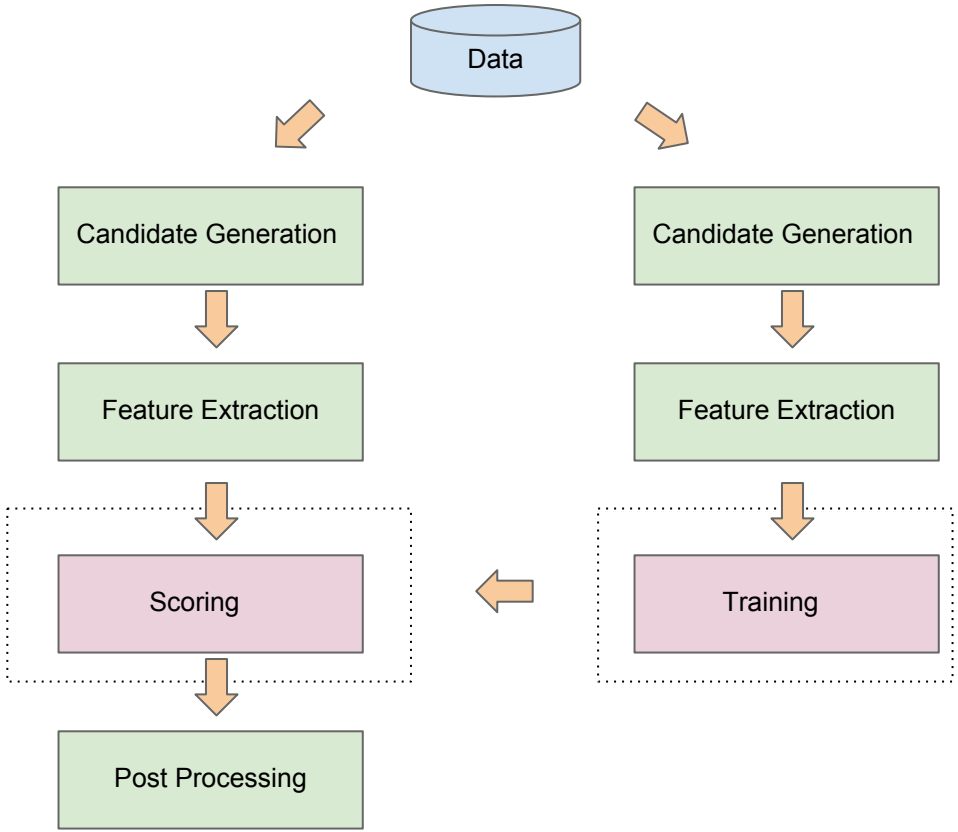
- External platforms at best can deploy “predictive models”, as services, not end-to-end online systems
- Gains come from optimizing the whole pipeline, not just algorithms.
- Latency: tens of milliseconds. Managing sharding, batching, data locality, caching, streaming, stragglers, graceful degradation...
- Real world systems -- boosts, diversity constraints, holes in data, skipping stages, hard filters... sounds familiar?



Reason # 5

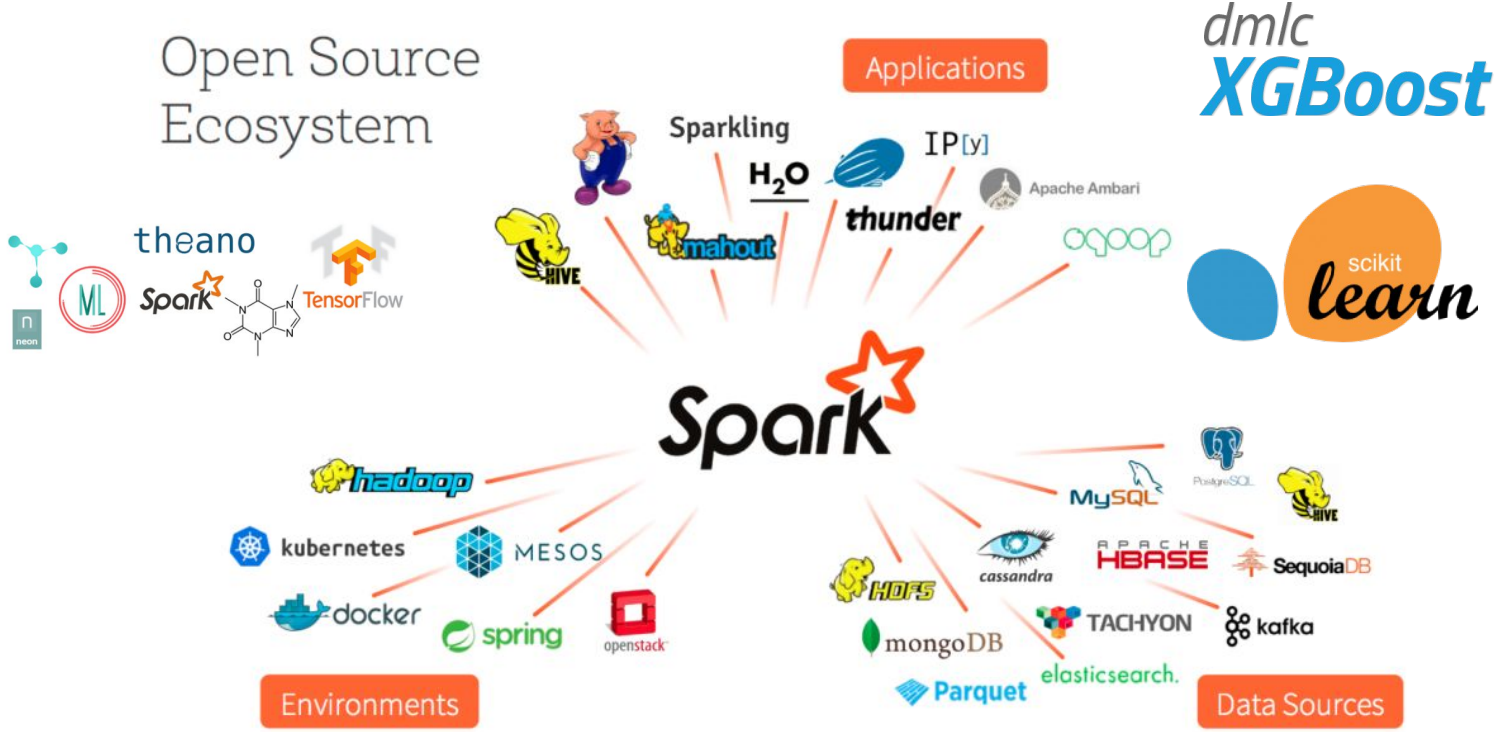
Blurry Line Between Experimentation & Production

- We want the same code/systems/tools to work for both experimentation & production.
- But we need to carefully “control” the production code to keep it be fast.
- So need to “control” offline experimentation systems too.



Reason # 4

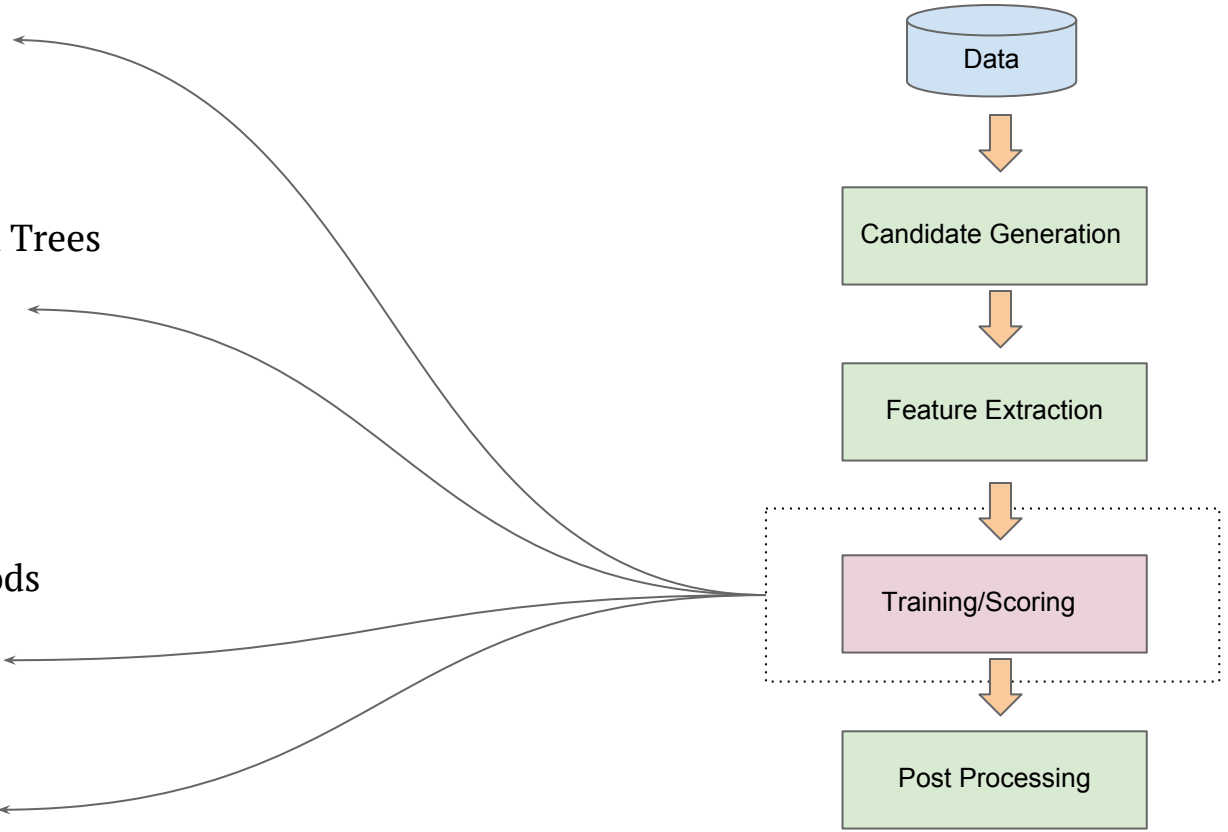
Openly Using Open Source


Open Source Ecosystem
Applications
dmlc
XGBoost
Sparkling
H2O
IP[y]
Apache Ambari
thunder
mahout
HIVE
theano
Spark
TensorFlow
scikit
learn
Spark
hadoop
kubernetes
MESOS
docker
spring
openstack
Environments
MySQL
PostgreSQL
HIVE
cassandra
APACHE HBASE
SequoiaDB
HDFS
TACHYON
kafka
mongoDB
elasticsearch.
Parquet
Data Sources

Production ML Algorithms At Quora

Quora

- Logistic Regression
- Elastic Nets
- Random Forests
- Gradient Boosted Decision Trees
- Matrix Factorization
- (Deep) Neural Networks
- LambdaMart
- Clustering
- Random walk based methods
- Word Embeddings
- LDA
- ...



- Open source is great -- lots of great technologies!
- Commerical ML platforms are also open sourcing stuff.
- Learning and cherry-picking favorite parts from ANY open source systems.
- May write our own algorithms too (e.g QMF)
- Building own platform = controlling the delegation, not lack of delegation

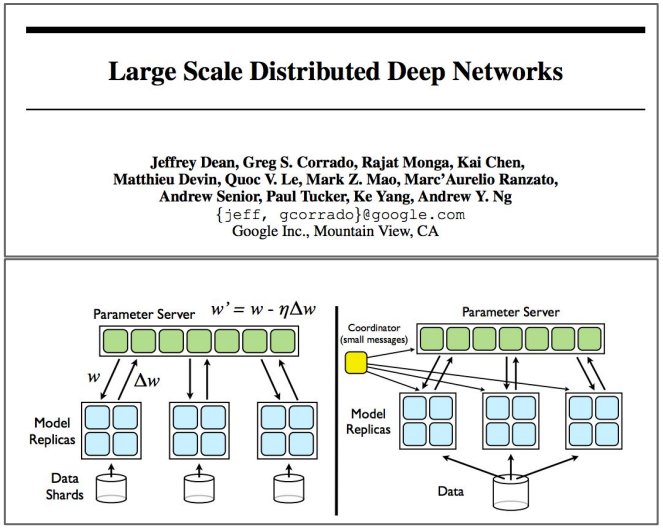
Reason # 3

Commercial Platforms’ Offerings Are Not Super Valuable To Us

- Main offerings of external platforms are:
  - Lower operational overhead of running machines
  - Out-of-box distributed training.

- Operational overhead
  - Gets amortized over time
  - Shared with non-ML infrastructure.

- Can often train most models in a single multi-core machine.



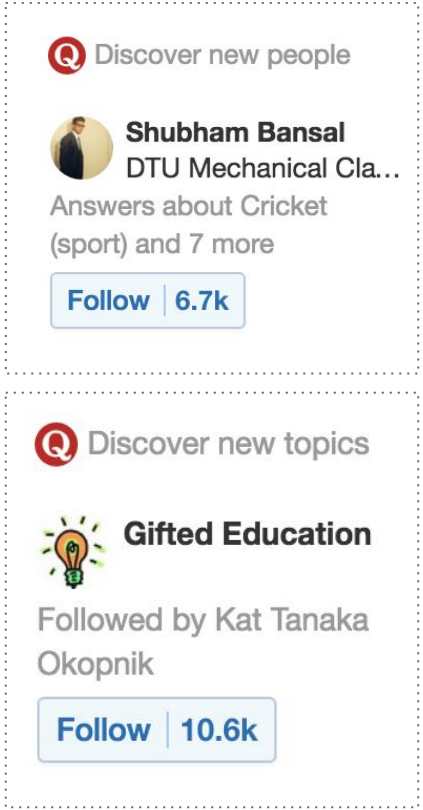
Reason # 2

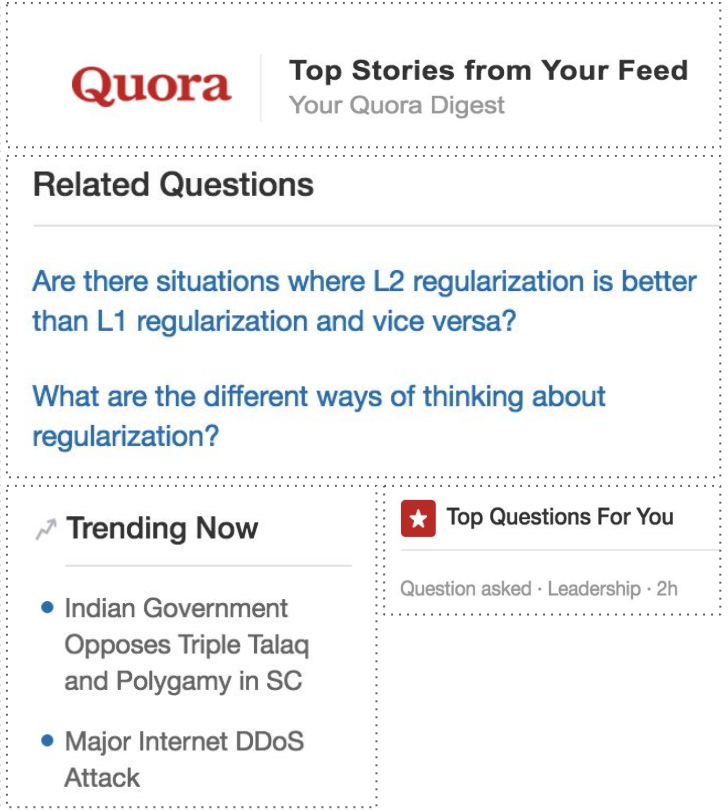
Blurry Line Between ML & Product Dev

Blurry Line Between ML/Non-ML Product

Quora

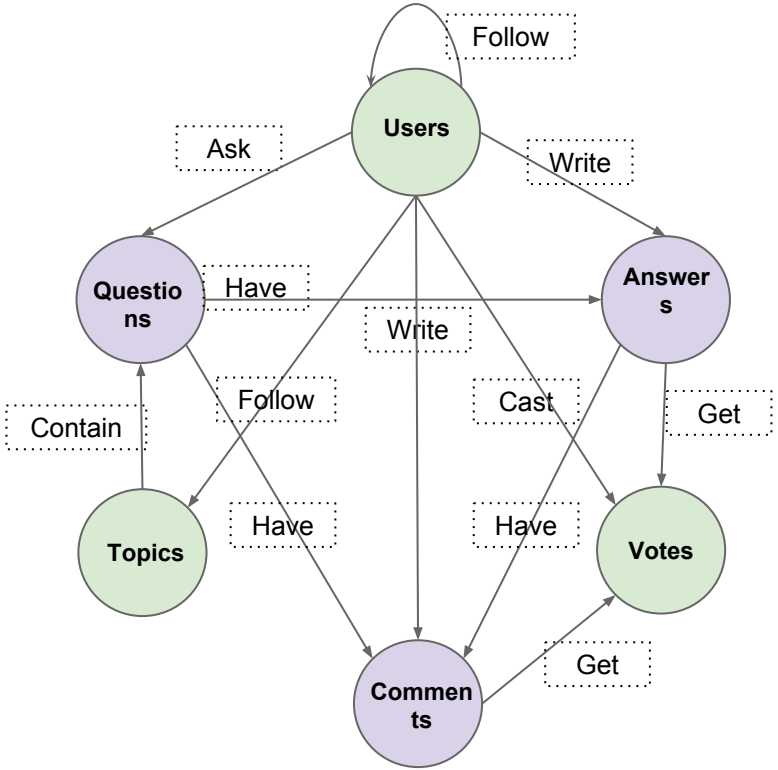
- Answer ranking
- Feed ranking
- Search ranking
- User recommendations
- Topic recommendations
- Duplicate questions
- Email Digest
- Request Answers
- Trending now
- Topic expertise prediction
- Spam, abuse detection
- ....

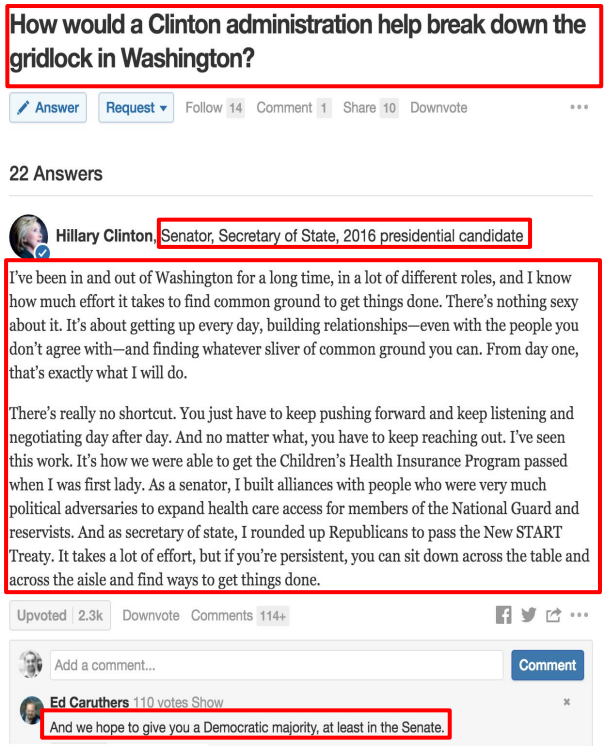




Blurry Line Between ML/Non-ML Data

Quora





Billions of relationships and words

Blurry Line Between ML/Non-ML Codebase

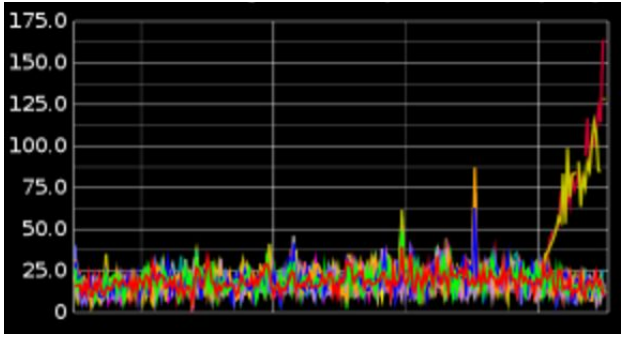
Quora

- Integration with other utility libraries/services
  e.g A/B testing, debug tools, monitoring, alerting, data transfer, ...

- Empowering all product engineers to do ML.

Variant	
True	Control
1,358,942	1,813,943

% CHANGE:
+1.9%
95% CI:
(+0.8%, +3.0%)



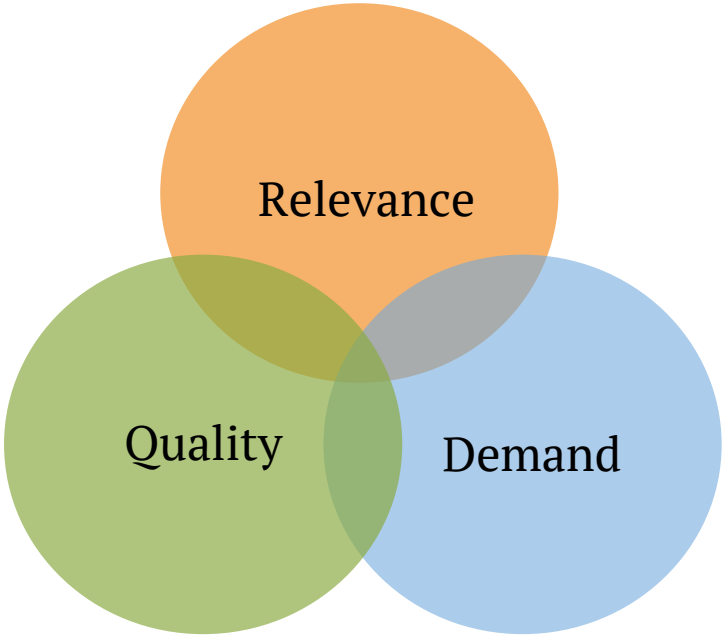
Reason # 1

ML As Quora's Core Competency

ML: Critical For Our Strategic Focus

Quora

- ML gives us a strategic competitive advantage.
- Want to control and develop deep expertise in the whole stack.
- Quora has a long term focus -- investment in platform more than pays off in the long term.
- Single most important reason to build ML Platform!



Summary

- Anyone doing non-trivial ML needs an ML platform to sustain innovation at scale.

- Build vs buy decision is not all-or-nothing.

- Surface area and importance of ML are deciding factors in the build vs buy decision.

Thank You!

YES, WE ARE HIRING :)

Nikhil Garg
@nikhilgarg28