

A Machine Learning Journey

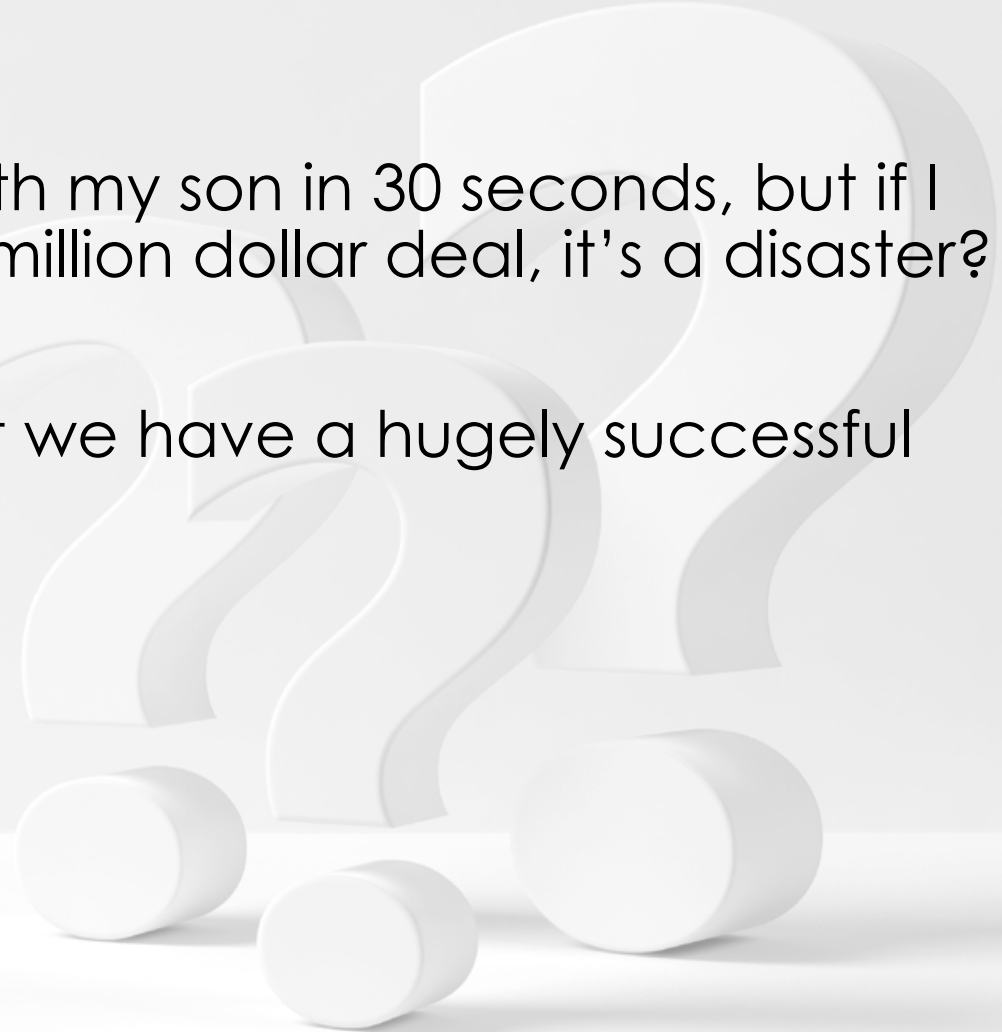
June 5, 2019

Oliver Sharp, Chief Solution Architect



The Two Fundamental Questions

1. Why can I find a viral pet video with my son in 30 seconds, but if I need a white paper to drive a 50 million dollar deal, it's a disaster?
2. If we solve that problem, shouldn't we have a hugely successful business?



The Obvious Answers

1. Machine learning! Duh.
2. Of course!



The ~~Obvious~~ Actual Answers

1. Machine learning! ~~Duh.~~

Well, yes .. But. Google knows a thing or two about search. Yet the Google Search Appliance didn't do a great job solving the enterprise search problem. They discontinued it.

2. ~~Of course!~~

Well ...

Getting Them Off the Couch

We built a solution. Now you can find the white paper as quickly as a pet video. Who could resist?



Well, a lot of people, it turned out

“I have other problems that are more important”

“That's something IT deals with”

Highspot Customers



Start From The Pebble in Their Shoe

Sales Content Management

Content Stored in Many Places

Content Difficult to Organize

Sales Can't Find Content

Sales Uses Old Versions

Sales Pitching

Email Sent into Black Hole

Live Pitching Has Friction

No Holistic Pitch Dashboard

Sales Doesn't Update CRM

Marketing and Sales Analytics

Measure Content Performance

Measure Pitch Performance

Track Changes by Sales

Identify Best Practices

What it Takes to Make Customers Successful

- 
- A person with their back to the camera, arms raised in a victory pose, silhouetted against a bright sunset over a city skyline. The sun is low on the horizon, creating a strong lens flare effect. The city skyline is visible in the background, with various skyscrapers and buildings. The overall mood is one of triumph and achievement.
1. Beautifully designed software ...
 2. With a touch of magic ...
 3. And a deep partnership

ML is the Magic

We've built

Classifiers

Clustering

Hierarchical agglomerative
complete-linkage

Collaborative filtering

Search ranking / NLP



Our tech stack includes

- SOLR/Lucene
- Clojure and Spark
- Algorithm libraries

So what have we learned?

ML is a Contact Sport

You don't reason your way to glory, you experiment your way there – this is Mythbusters, not physics class

If you don't have data, you don't know sh*t



The Quest for More and Better Data



There are 75 ways of doing anything ... and they all work

Logistic regression vs. multi-normal logistic regression vs. Probit regression vs. support vector machines vs. linear discriminant analysis vs.

More good clean data is **far** better than a fancier algorithm

Business Data is a Challenge



It is sparse and noisy

It is very easy to over-fit

Most of it is useless for solving real user problems

It is hard to know if you made things better

ML is an Approach

- You need to choose tools at the right level of abstraction
- You need a *workbench* so you can quickly iterate and try things
- Traditional criteria still matter - code quality, scalability, ...
- But so do new ones – flexible data formats, fluid mix and match
- *You cannot outsource insight* – there is no black box



ML isn't a component you bolt onto your application – you are as much *investigator* as *builder*

The ML journey is Never Finished



Your first ideas are always at least somewhat wrong

Build a learning environment that lets you get smarter over time

New insights and data constantly evolve your approach

Points of Leverage

You need to find where the insights from ML actually *matter*

Some were obvious .. and right

- Search
- Recommendations

Some seemed obvious .. but were wrong

- Optimized news filtering

Some were discoveries

- Content similarity
- Trending sort
- Salesforce ratings



Build with Glass Pipes

Traditional system architecture

Break it all down into abstract
little boxes for scale and flexibility

... is *terrible* for machine learning

You need the *whole* context and
the ability to modify it



Your (Enterprise) Big Data isn't that Big



You can fit a heck of a lot of data on a 4TB hard disk

It is **so** much faster to work on data that lives on one machine than to use Hadoop on a server cluster

To Summarize ...

ML is magic ... but most business problems aren't solved by (just) an algorithm, no matter how good it is

To liberate that magic, you need

- The right customer problem

- The right data

- The leverage points

- The right tools and algorithms

... and a commitment to constant learning and refinement

