



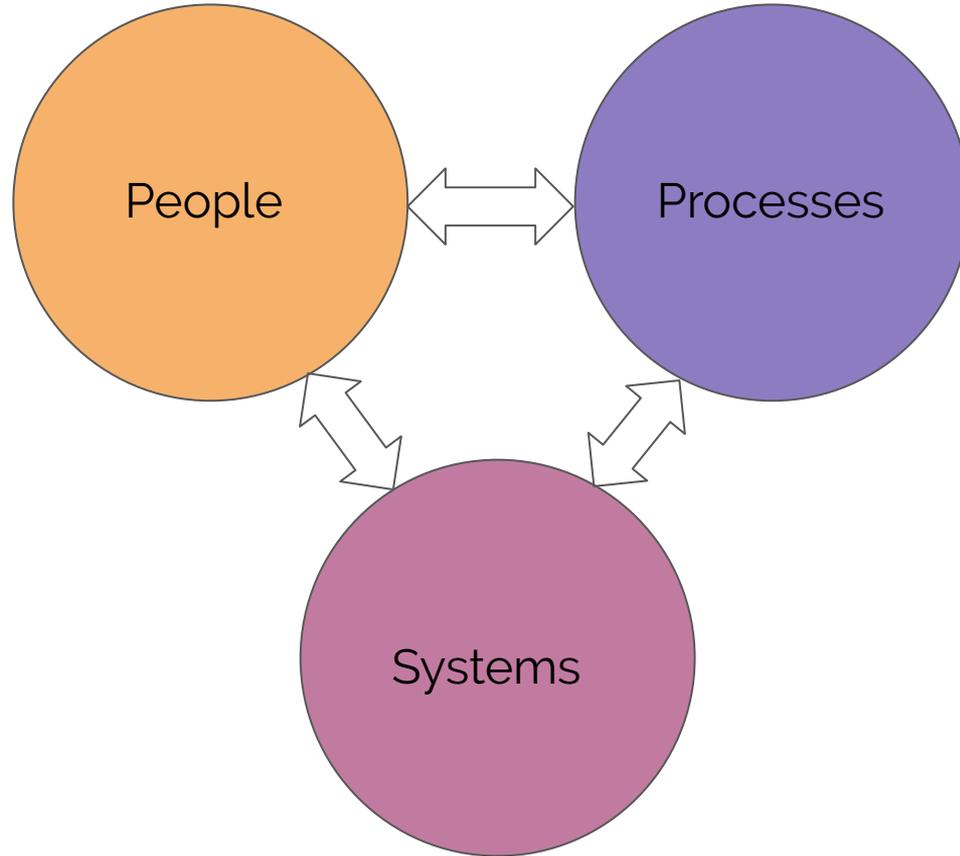
May 30, 2019

Machine Learning from System Quality Perspective

Mikhail Iljin

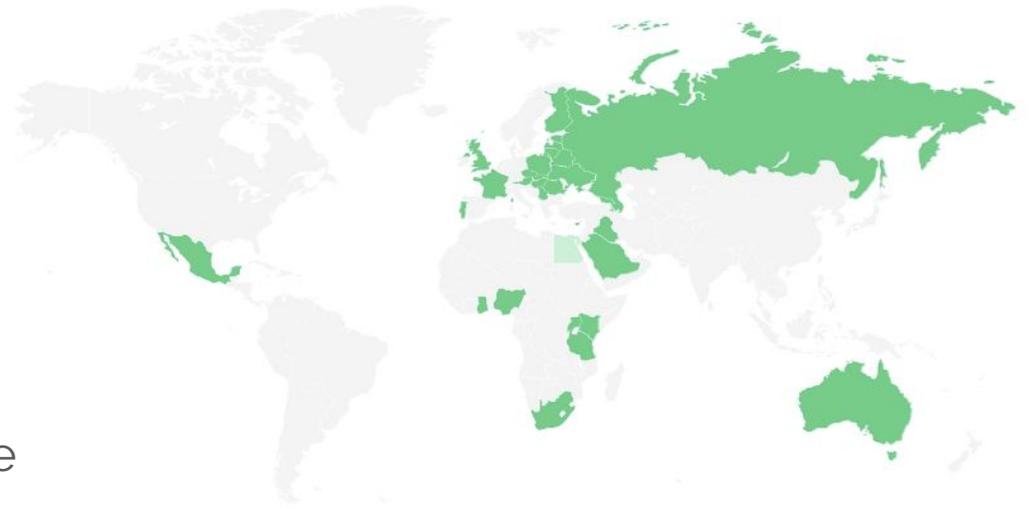


What is scaling data science?



There is a company named Bolt

- In the last 2 years
 - 20x more rides
 - 15 -> 30+ countries
- How to optimize all this?
 - Data science to the rescue



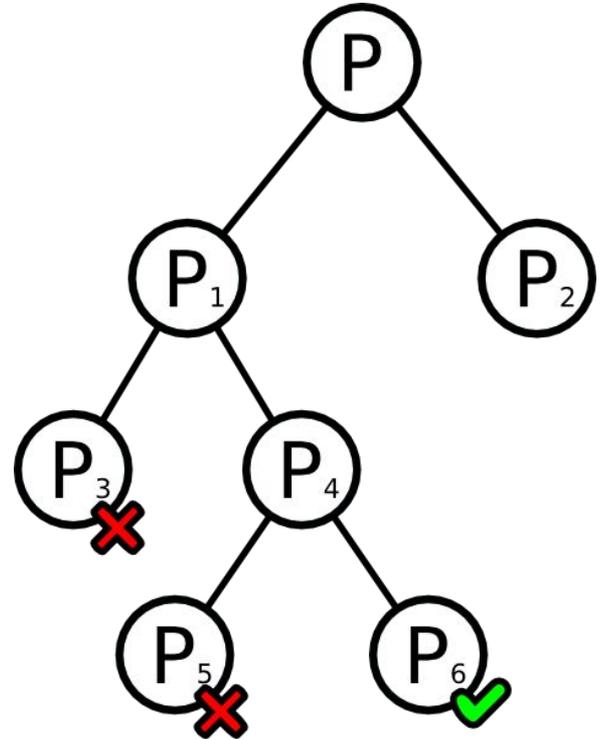
What is it?

- Pieces of software
 - AWS, NodeJS microservices, Android and iPhone apps, MySQL
- It needed QA
 - Unit, component, integration, UI tests

Let's build a data-driven company in 30 minutes!

What happens to software when ML is added?

- Line between *code* and *data* is blurred
- Logic is non-deterministic
- Logic is very branched and opaque
- “Quality” is not absolute, depends on context
- Influence of ML components overlap



What happens to software when ML is added?

- You have to be suspicious and vigilant about your work
- Everyone must be ~~a bit crazy~~ passionate to think, talk and figure out what's next

Software changes itself by itself

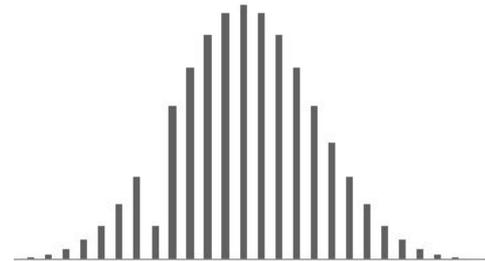
A man with dark hair, wearing a white button-down shirt and a dark jacket, stands in a blue, ethereal environment. The background is filled with soft, glowing light patterns. A bright, glowing white ball of light is visible in the lower-left quadrant of the image. The overall mood is serene and healing.

Imagine your pain as a white ball of healing light

When do you (we) need data science?

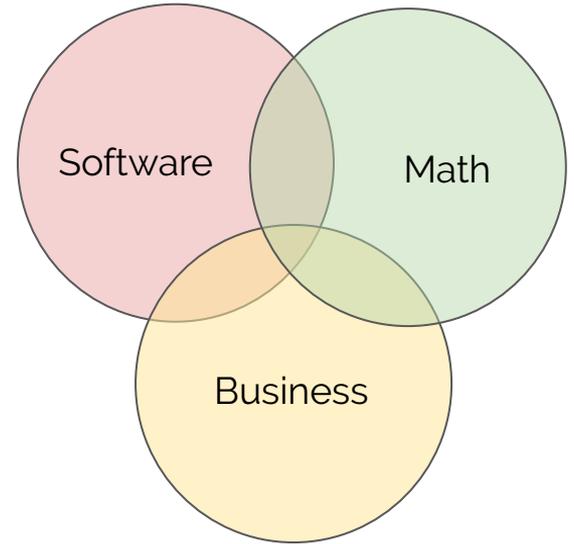
- Software must behave quantitatively (based on data and statistics)
- Settings should adjust quantitatively
- Get last several % of efficiency
- Figure out what's going on in your data

```
if (a != 3 || b != 4)
{
  //do something else
}
else
{
  //do something
}
```



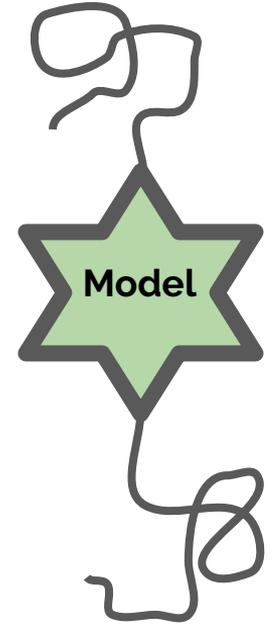
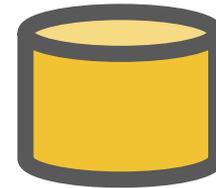
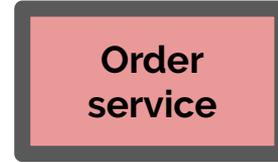
From exploration to prototypes

- Business - *do you really understand the problem?*
- Data quality - *does the data represent the world?*
- Data leakage - *will it still work in live?*
- Model - *start with simple baseline*



From prototypes to live deployment

- Data scientist's code **is not** production grade code
- Communication with developers
 - Understand each other ❤️
- Is this deployable?
 - Data available in live
 - Latency of prediction
 - Integration points with backend - *no hacks*



From live to being sure it's better than before

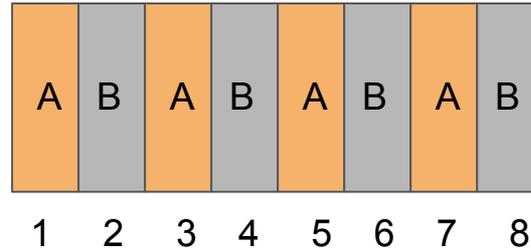
- Right metrics to test
- Right type of test
- True randomization

	Delta 	% change 	P-Value 
Price (eur) (105254/1002461)	-0.042±0.034	-0.794	0.015
Recalculation rate (105254/1002461)	0.001±0.002	0.593	0.566

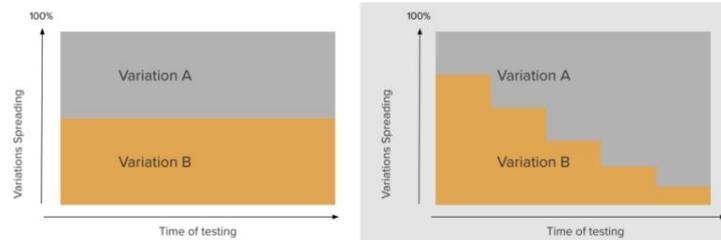
- 2% difference in means **is not** 2% difference in money
- Statistical significance is not enough - know the context
- A lot of tests = some will be statistically significant by chance
 - P-value **0.05**, **50** tests, **92%** that at least 1 is by chance
- Long-term effects also matter

From live to being sure it's better than before

- Interleaved tests



- Bandit tests



- Simulator

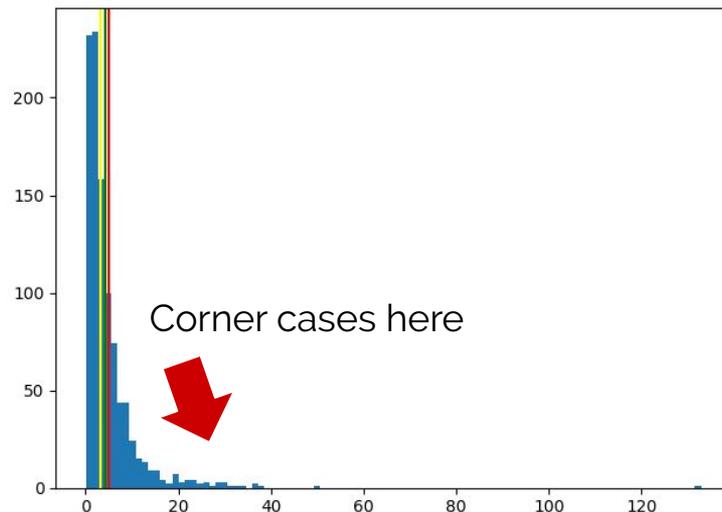


From live to live issues!

- Have fallbacks for corner cases
- World may change suddenly

What is good monitoring?

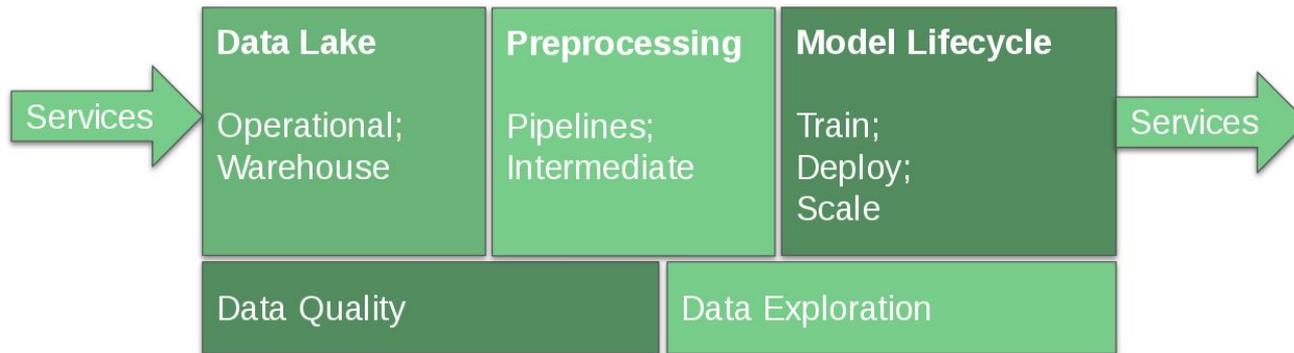
- People don't ask you, but share your graphs with each other
- You know of issues before someone told you
 - Or actually before they happen



From live issues to constant maintenance

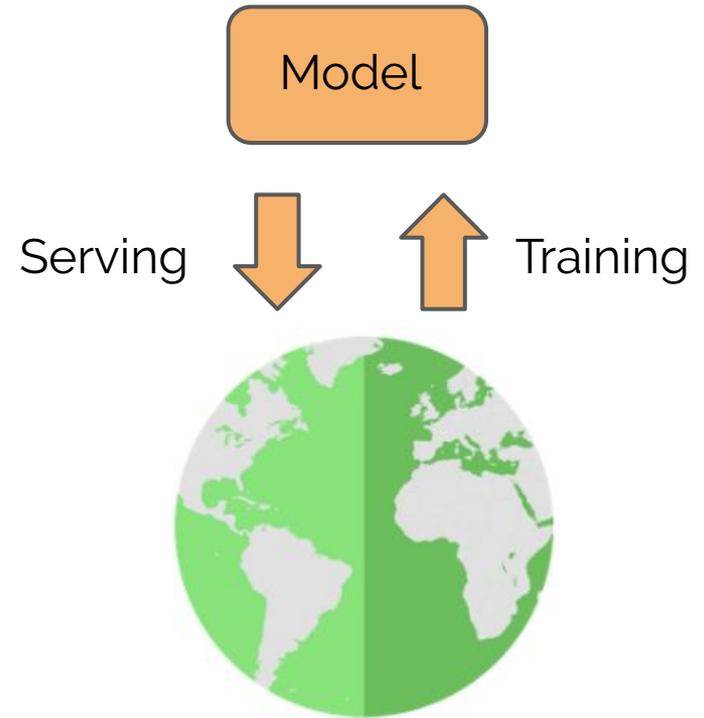
- Manual work = human error
- Data scientist's time is too valuable for maintenance
- We need control and flexibility

Data Science Platform

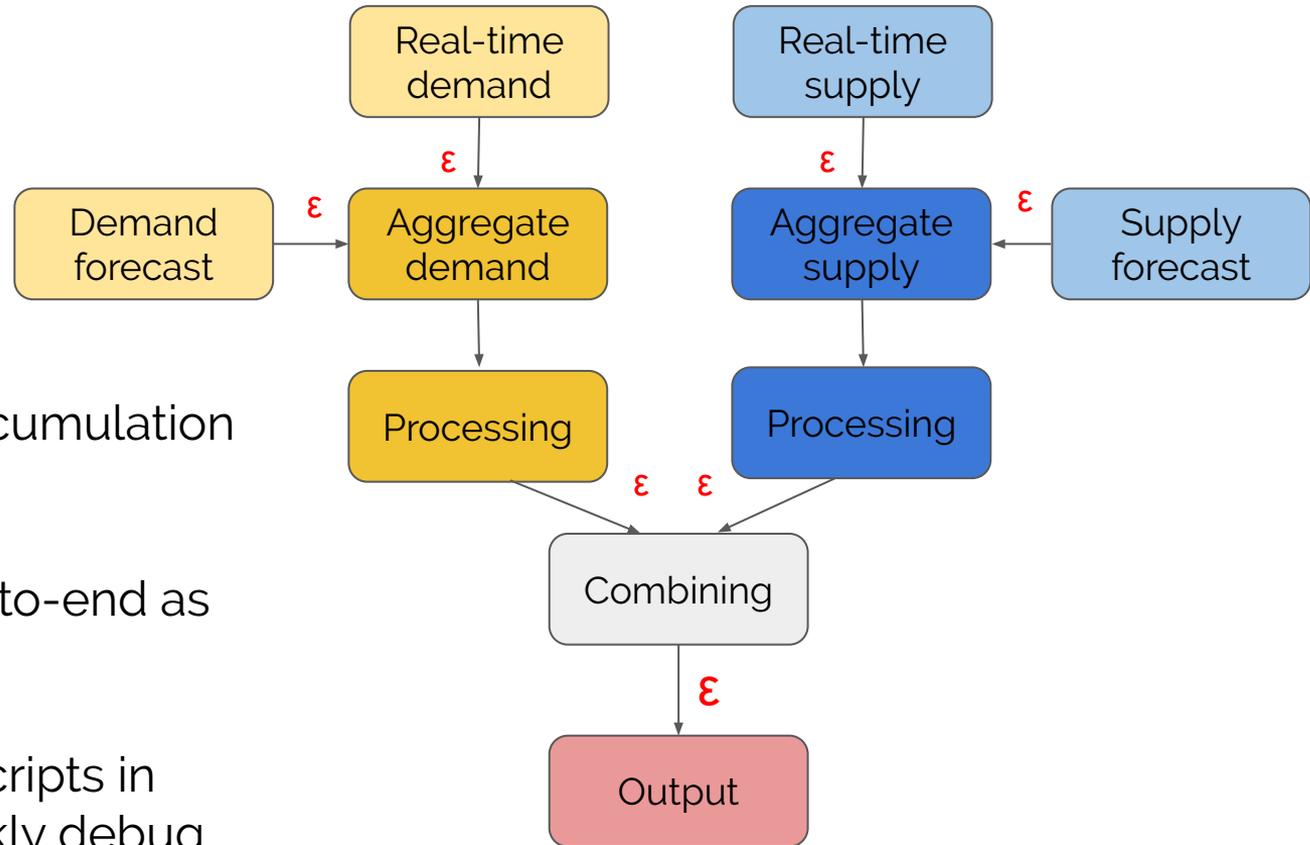


From our own work to outside world

- Model has inputs from other systems
 - Do you have them covered and monitored?
- Models may influence the world
 - Do you account for that?
- World-model feedback loops
 - Model changes the world, then learns the change, then...



From one-model to many-model components

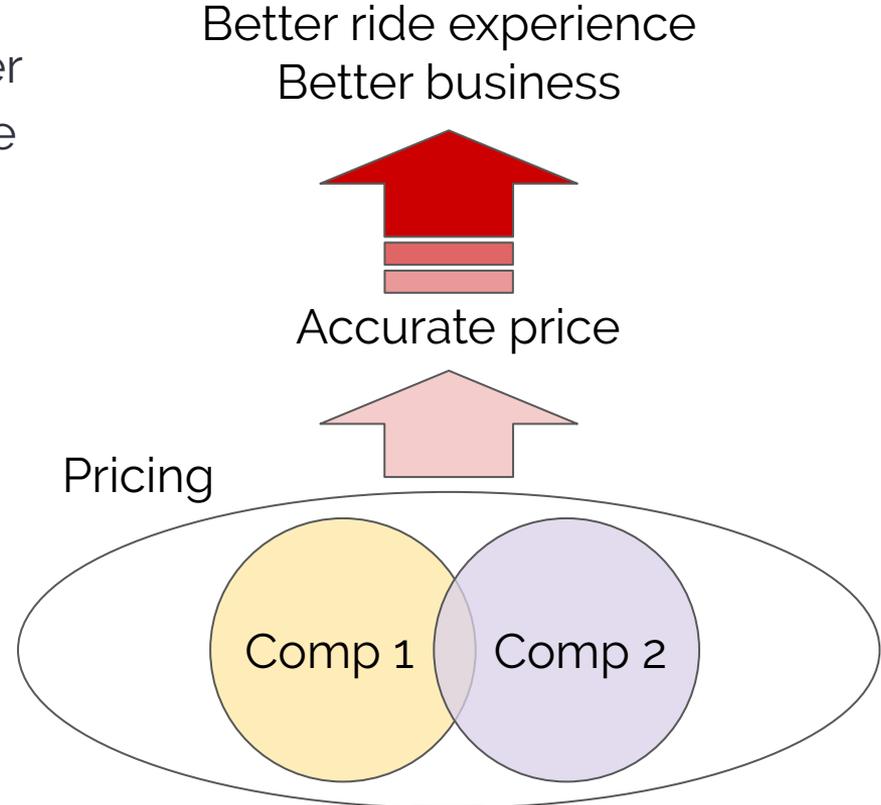


- Beware of accumulation of errors
- Make as end-to-end as possible
- Have tools/scripts in place to quickly debug

From separate components to architecture

- Components influence each other
- Components working towards the same goal

- **Everything** is actually working for the same goal
 - Do you have a theory for how everything works together?





```
01001001001111011101001001001111
11011101001001001111011101001001
01001111011101001001001111011101
01001001001111011101001001001111
11011101001001001111011101001001
01001111011101001001001111011101
01001001001111011101001001111011101
01001001001111011101001001001111
11011101001001001111011101001001
01001111011101001001001111011101
01001001001111011101001001001111
11011101001001001111011101001001
01001001001111011101001001001111
11011101001001001111011101001001
01001001001111011101001001001111
11011101001001001111011101001001
01001001001111011101001001001111
11011101001001001111011101001001
01001001001111011101001001001111
11011101001001001111011101001001
01001001001111011101001001001111
11011101001001001111011101001001
```

