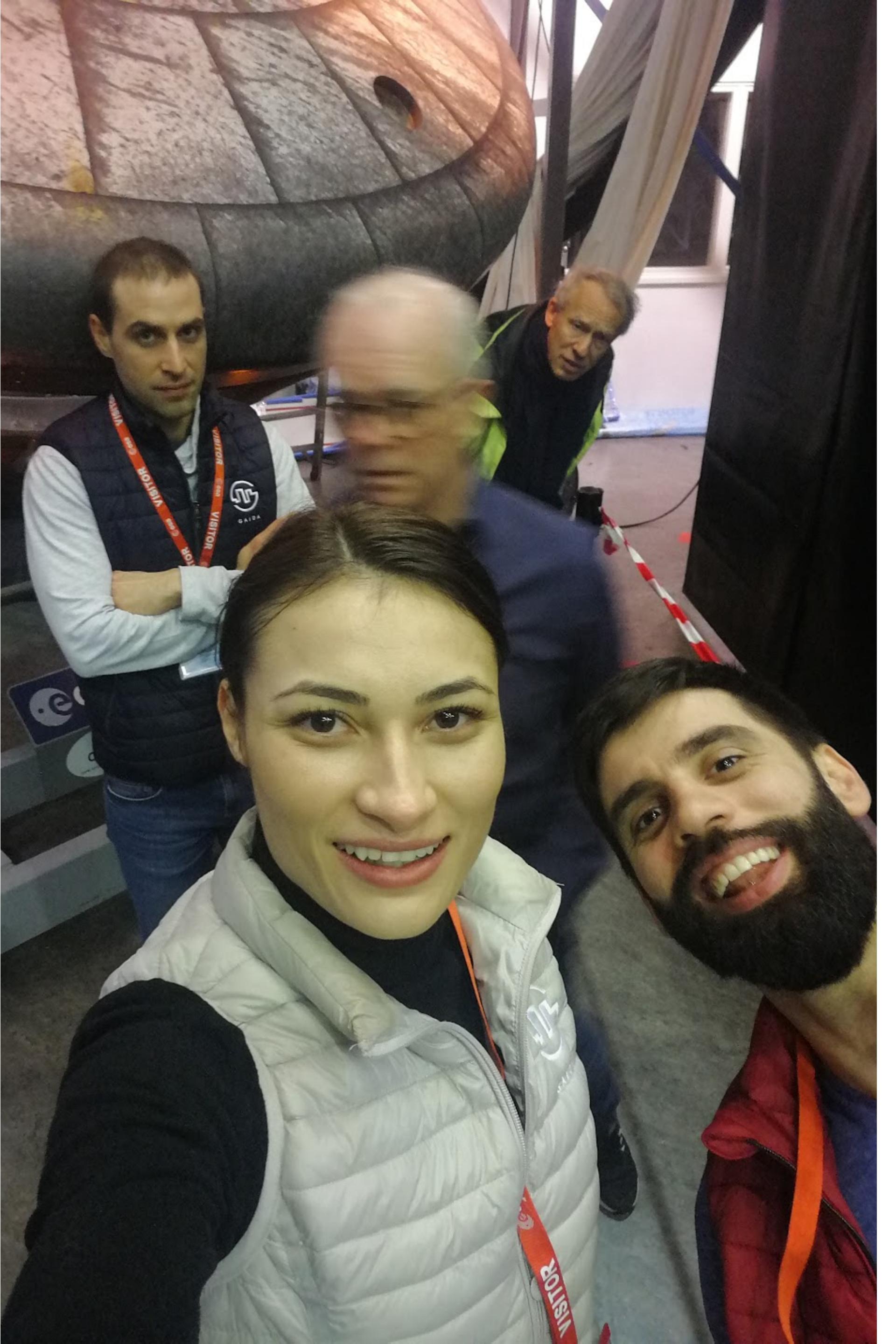


ML in PropTech

-

Concept to Production





Hello, who dis?



- Established 2018
- Milestones
- Future

Boyan

- Dev
- Data Eng @ scale
- FP monkey

Demir

- XP with SL, now into RL
- Armed bandit
- Afraid of commit(ment)

A photograph of a dark grey pigeon perched on a textured red pillar. The pigeon is facing right, looking towards a blurred cityscape with buildings, a road, and a car in the background.

The Story Today

- Every adventure starts with a leap of faith
- Fun with pipelines
- A look inside the data
- Recommender Systems tl;dr
- How to solve it
- How to ship it
- Results
- Lessons learned



Real Estate Inside/Out

- Old business, old tech
- Riding a giant wave
- Resistance to using tools

**...but the customer
is changing**

- Need information abundancy
- Expect being served 24/7
- It's all about the me me me

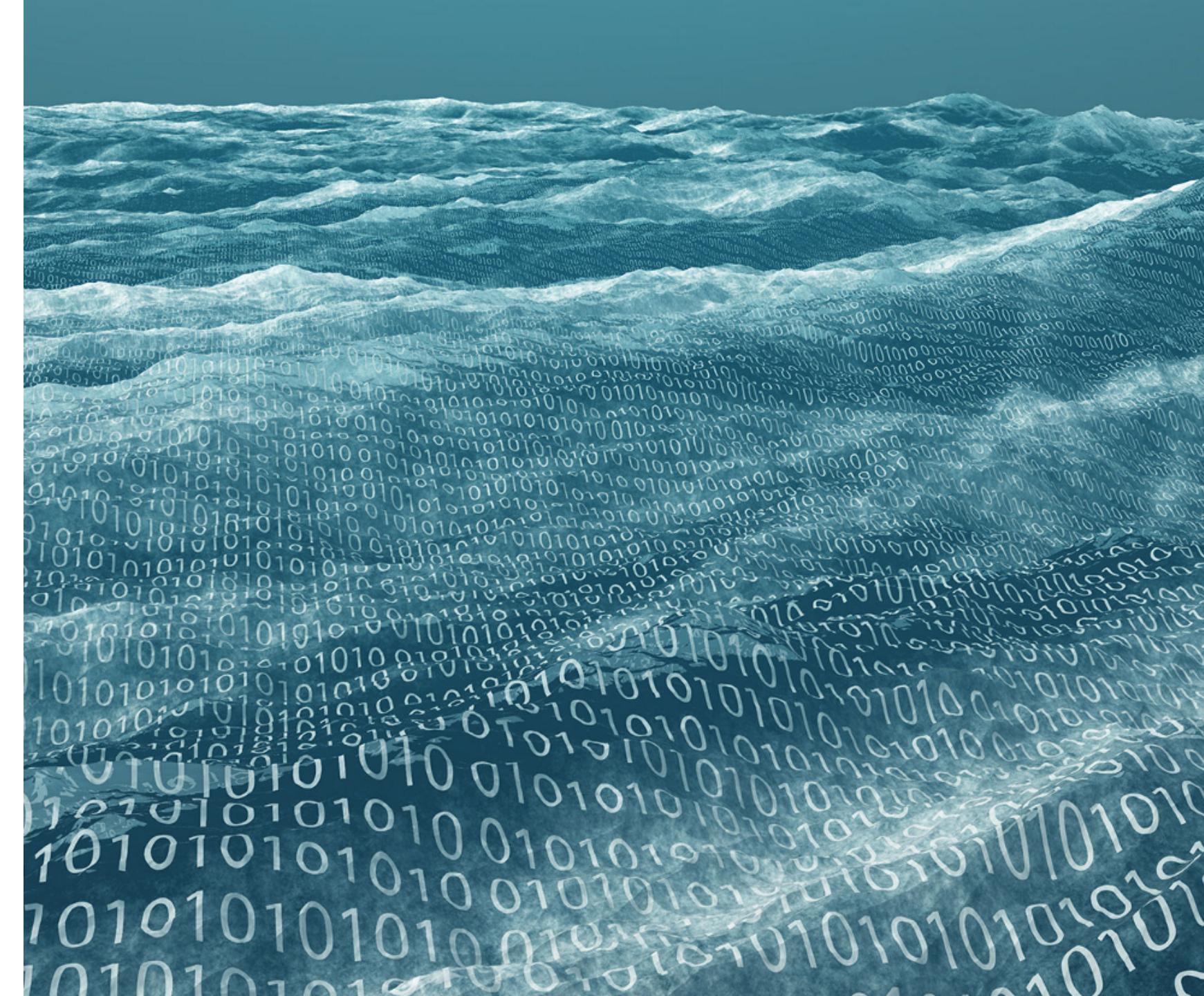




**...goes through
a phase of magic...**

(Clojure/Kafka)

**It all starts with
dubious design...**
(SQL/Normalization)



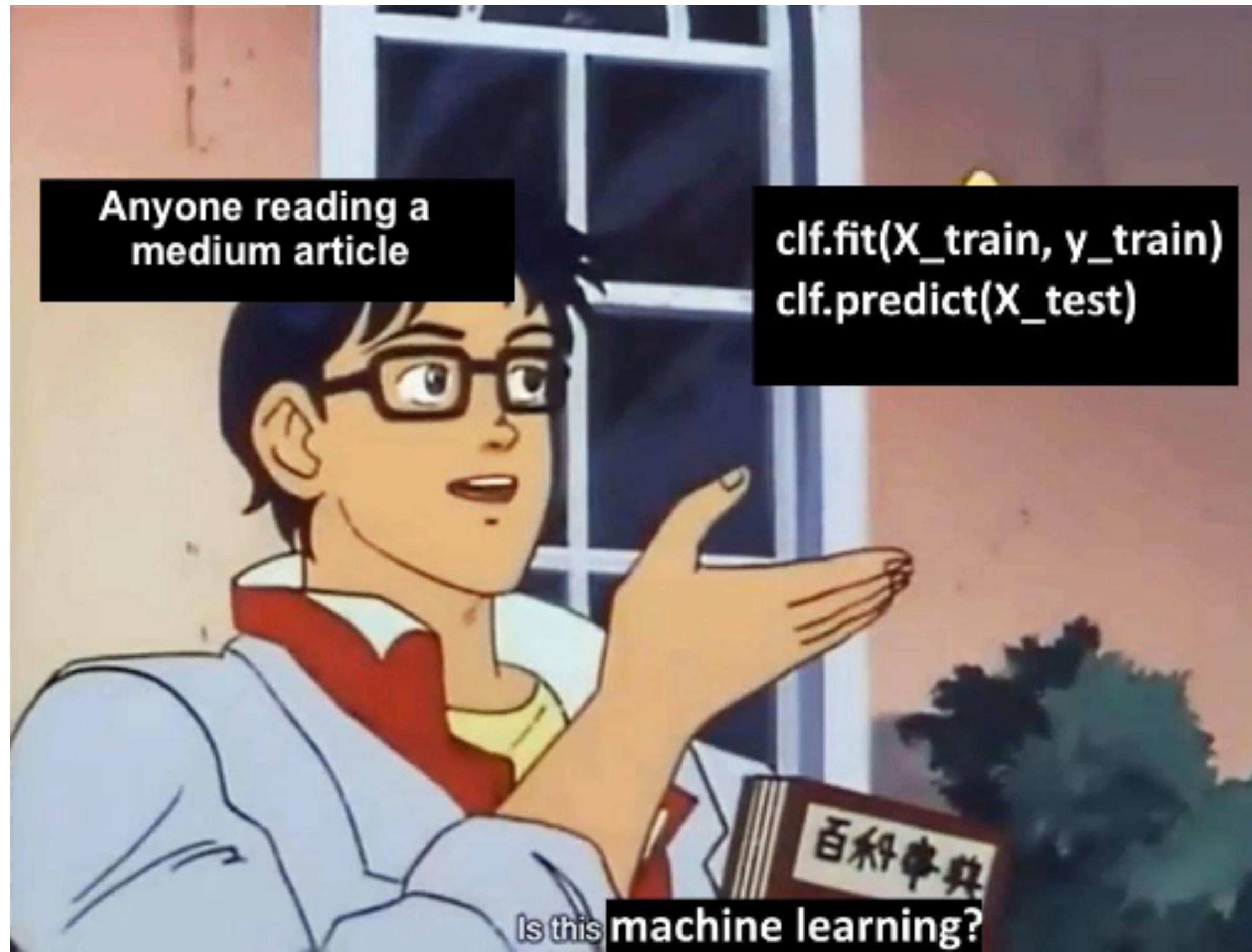
...into the data lake

Data Exploration

- Lack of culture around the importance of keeping a record
- Register the absolute minimum required
- Free input is the devil
- Single day journey
- Migration from obsolete systems without normalization



Meme slide goes here / It's time for MACHINE LEARNING*



*if on presentation - it cures cancer and fights aliens

How Netflix(Youtube) do(n't) it

CF

- Needs data. Lots of it
- Cold start problem
- Recs based on user behaviour



John	5	1	3	5
Tom	?	?	?	2
Alice	4	?	3	?

CB

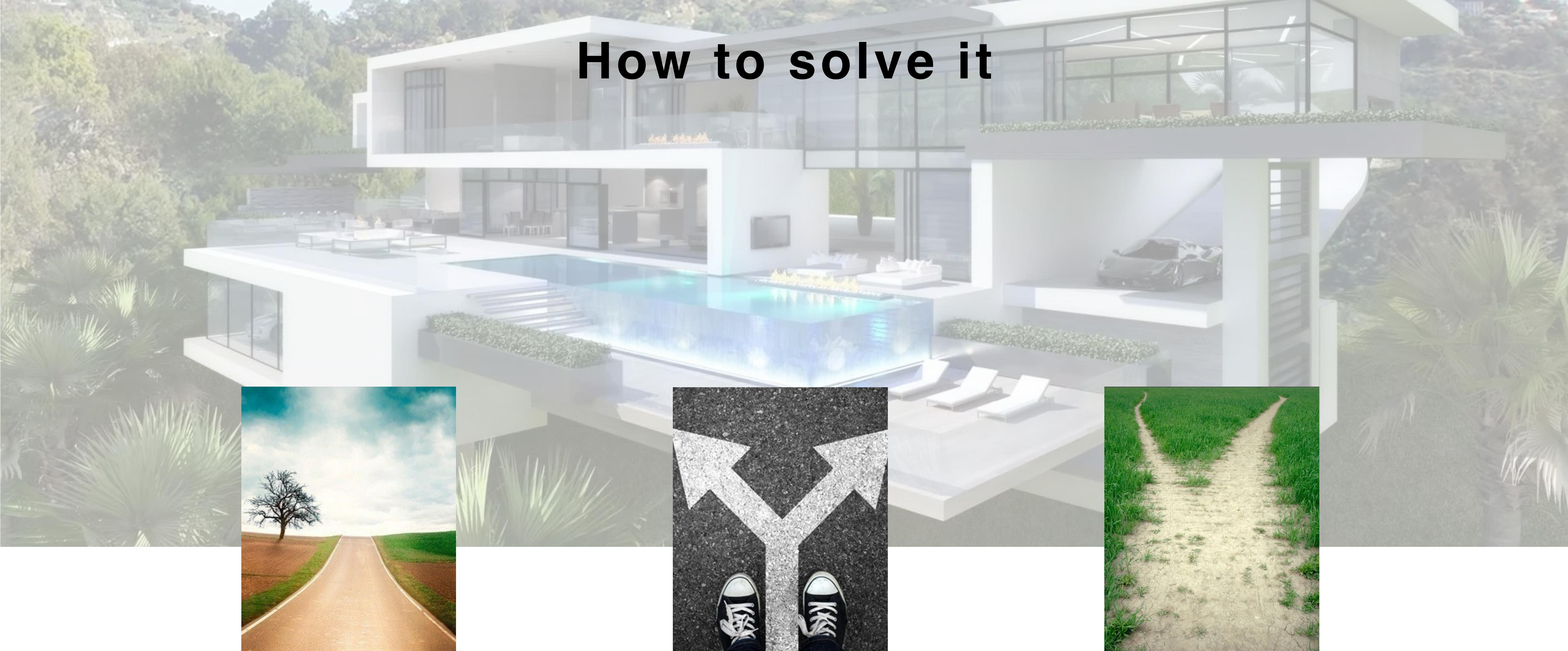
- Cold start for new users
- Feature based similarity
- Model per user, not a shared one

The Recommender we need, but not the one we deserve

- Constant influx of new users
- Very long repeat customer cycle
- Perishable items
- Limited interaction between user/items
- Buy/Rent are very different paths
- There is no such dataset(yet)



How to solve it



$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$



A dense, futuristic city street at night, possibly from the movie Blade Runner. The scene is filled with towering skyscrapers, numerous glowing neon signs in Chinese characters, and several flying cars suspended in the air. The overall atmosphere is dark and atmospheric.

Here We Discuss
The Future

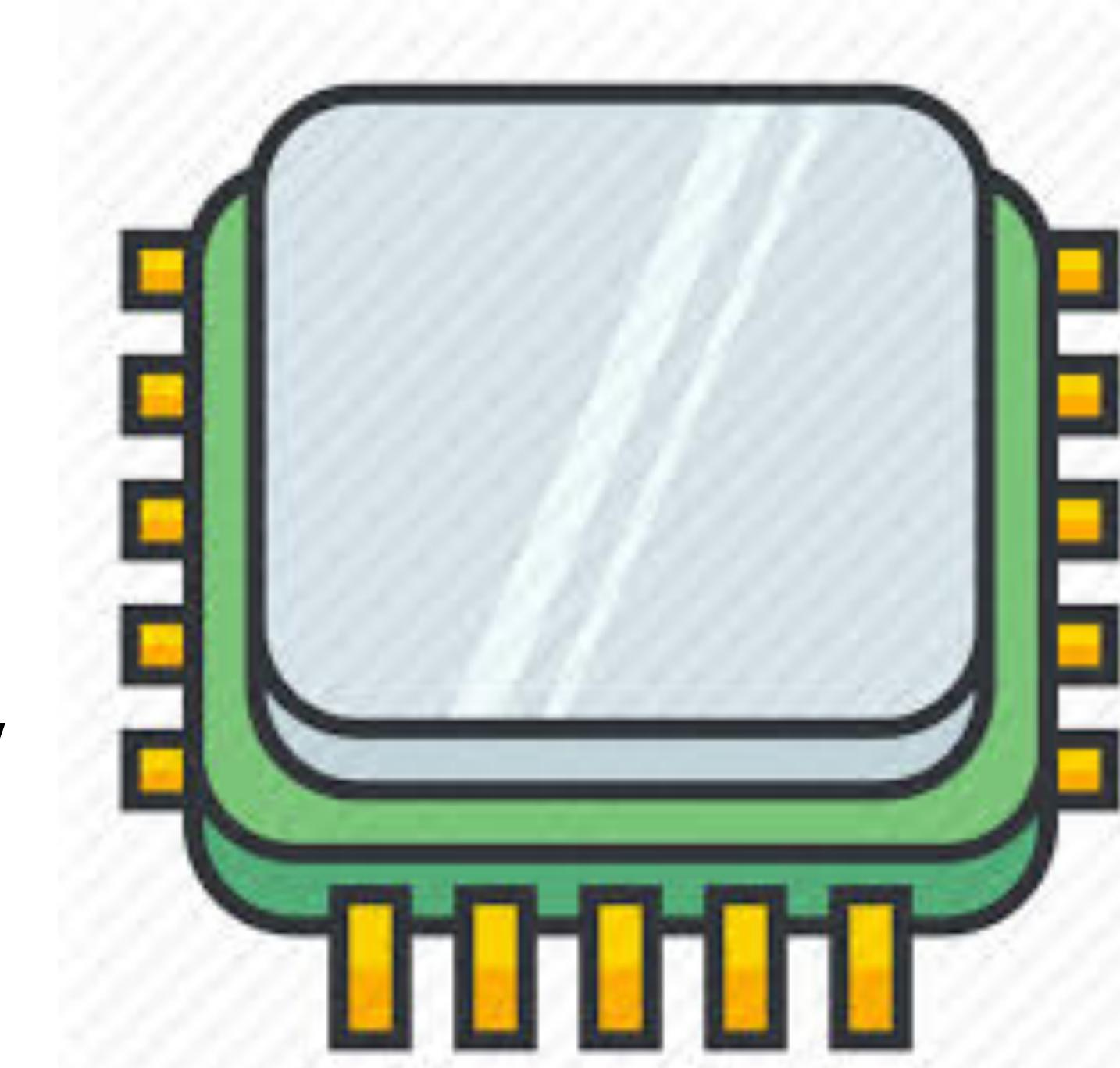
`__init__.py`
`core`
`serializers`
`simulation`
`train`
`transform`

Microlibs
Pattern
-
Notebooks to Library

How to ship it

ML Cores
(serialized)

HTTP
to the max



The Effect



- Served ~ 17000 customers, via 4 channels(mailing, sms, chatbot, mobile)
- ~35% open rate of communications with an average around 2.4 clicks per open
- ~16% re-activation rate (inquiry/viewing)

Lessons Learned

- Integrate high in the stack
- Data quality >>> Model complexity
- Business value does not need the most advanced tech
- Digitalization should ideally precede advanced ML
- Available tools will not always fit every specific problem
- It's hard being a fast company in a slow business world

HOW TO DO MATH:

1. WRITE DOWN
THE PROBLEM

2. CRY.

THE
END

BOYAN@GAIDA.AI + D.TONCHEV@GAIDA.AI