

An anatomy of scaling unsupervised learning for online gaming

Igal Grinis

A journey into machine learning architecture

What is Online Gaming?

Where does it hurt?

- High cost of user acquisition
- Regulation
- Need for near real-time response



What story data tells?

People text in

- typos
- intentional typos
- abbreviations
- emojis

19:44:08 ass hold
19:44:10 fck u
19:44:26 dealer card always 10 1st card
19:44:41 dammmm u
19:44:53 see
19:44:45 cheating
19:45:17 change dealer please



This is the third time today this user complains on fraud.

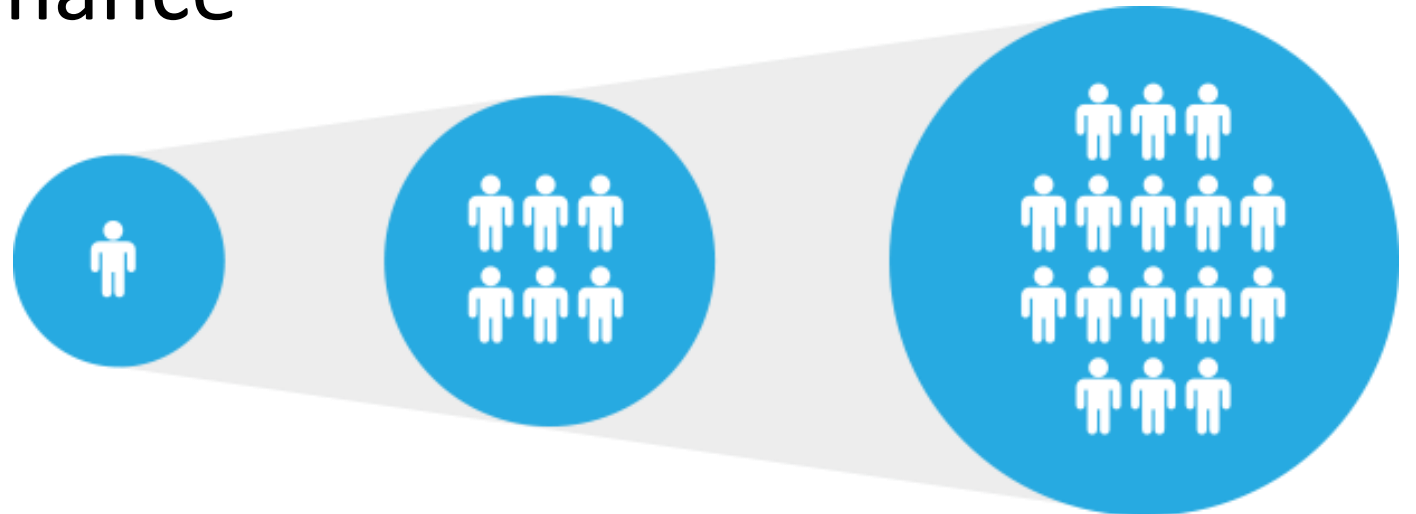
Choose your algorithm

- Keyword search (KWS)
- NLP Parsing
- Machine learning (ML):
 - Supervised learning
 - **Unsupervised learning**



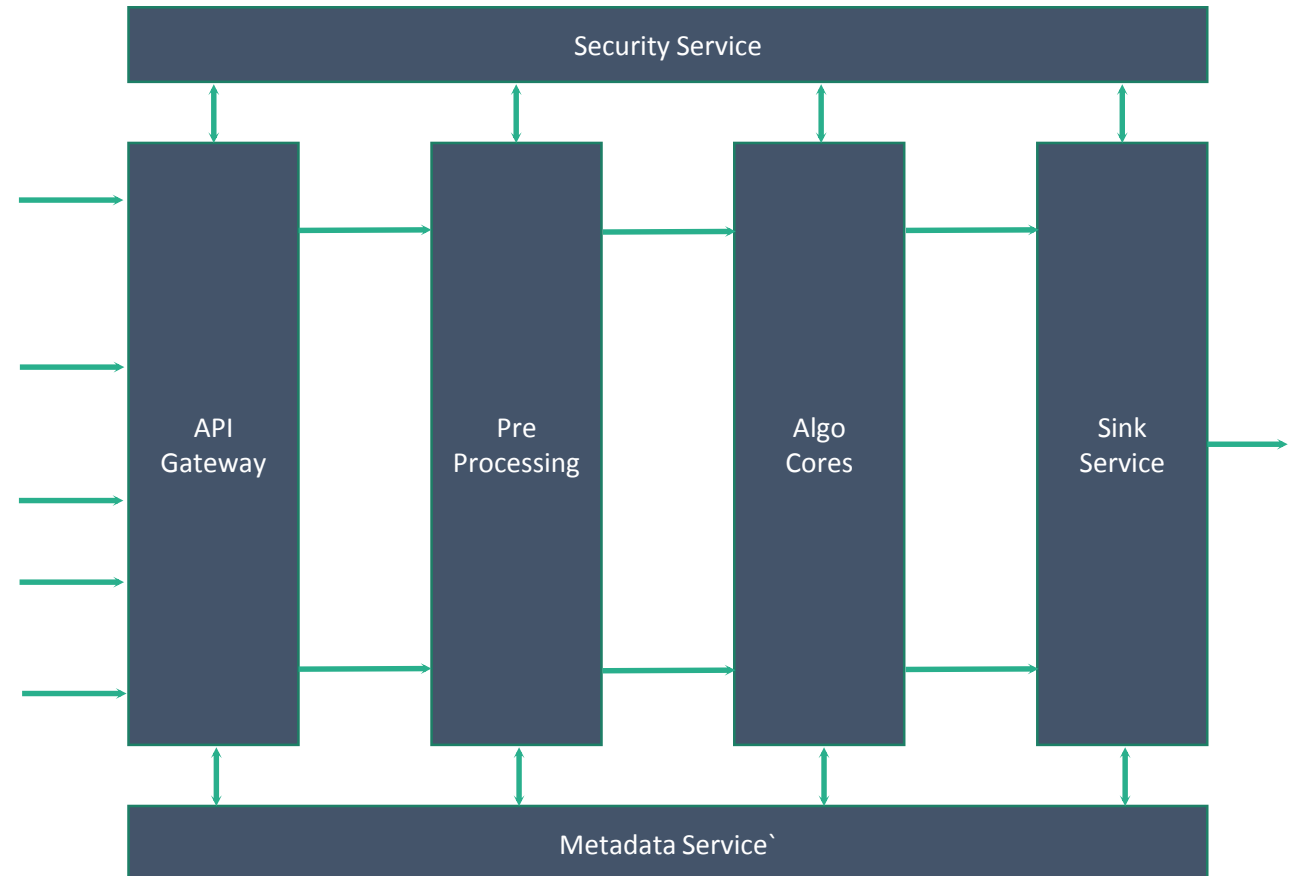
Small to Big Scale?

- Multiple customers/arenas
- Multiple languages
- Near real time performance
- Fast feedback loop
- Models flexibility

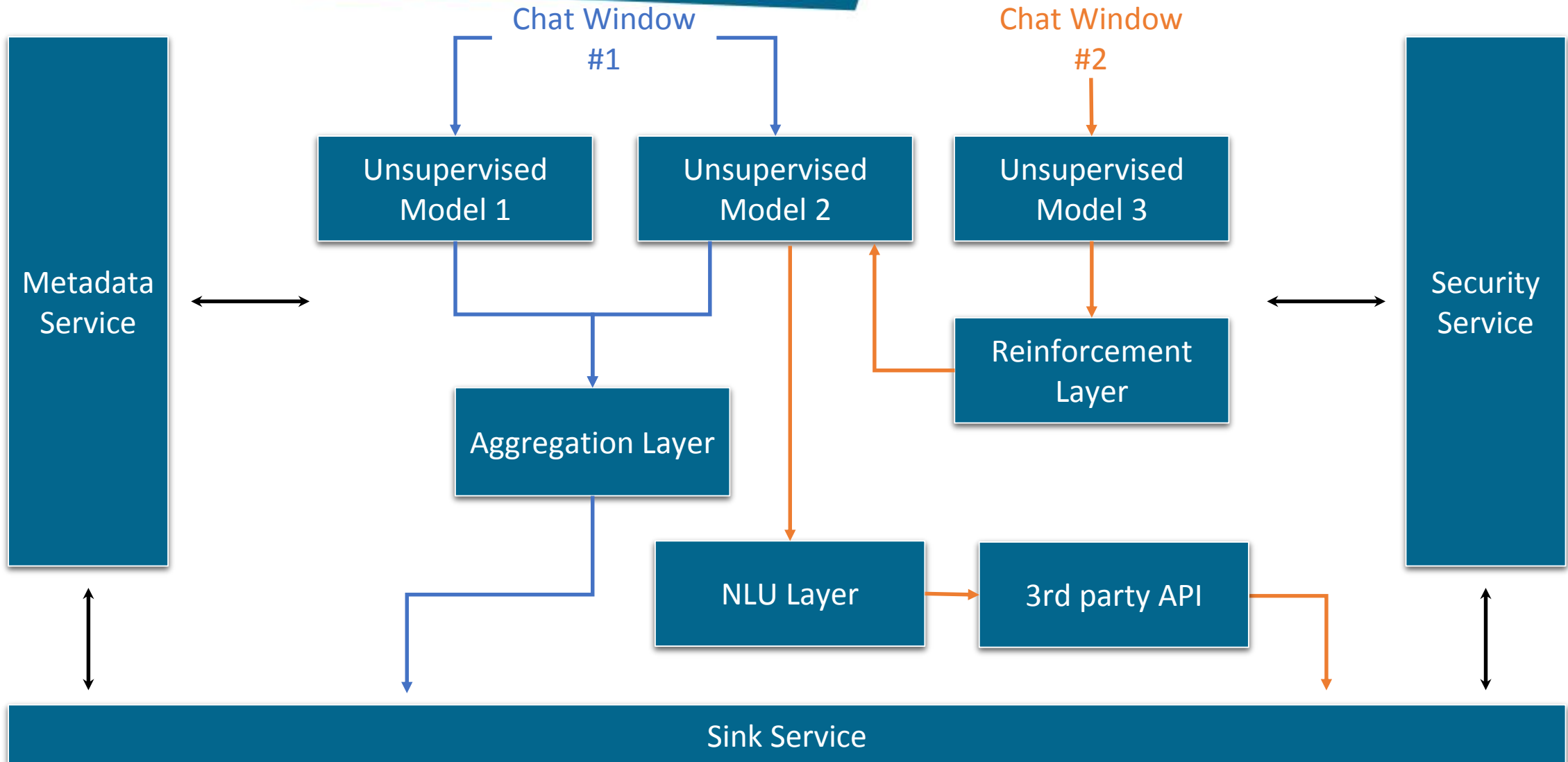


Serverless is the new Black

- Micro-services driven architecture
- Our ML pipeline is serverless
- Fully containerized platform
 - ✓ Docker
 - ✓ K8S orchestration
- AMQP Transport
- Autonomous scaling



Tomobox's Solution



Online vs. Aggregation

Texting language is rapidly changing

It's all about control...



Closing the loop...

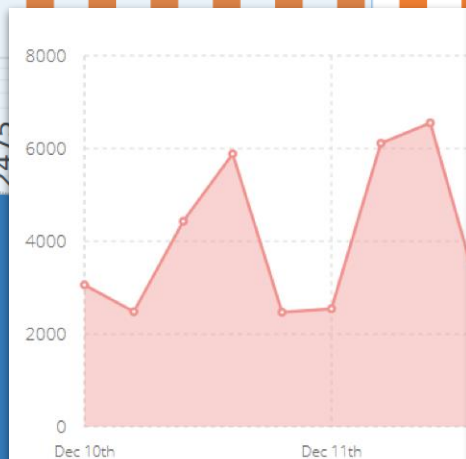
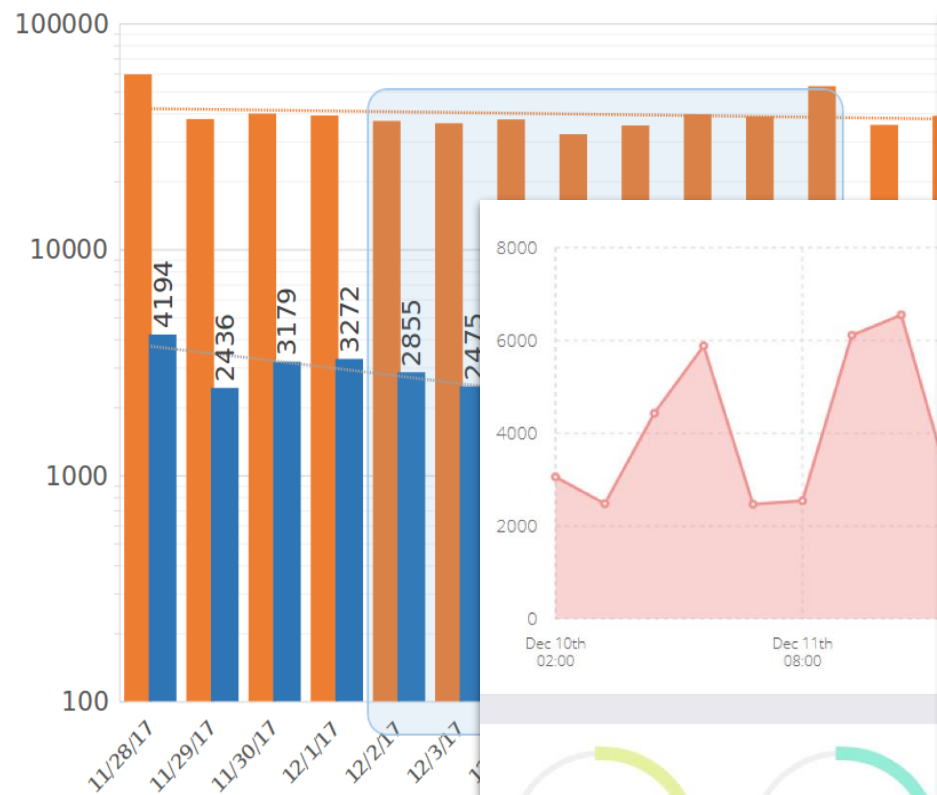
Are we done? No!

How do you test new models and compare them in an ever-changing environment?

Create a strong feedback loop!



Ever progressing

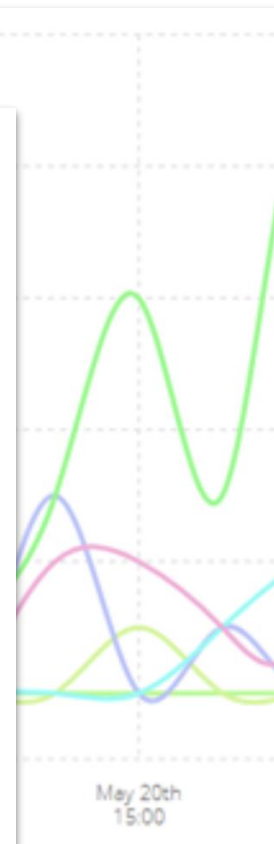
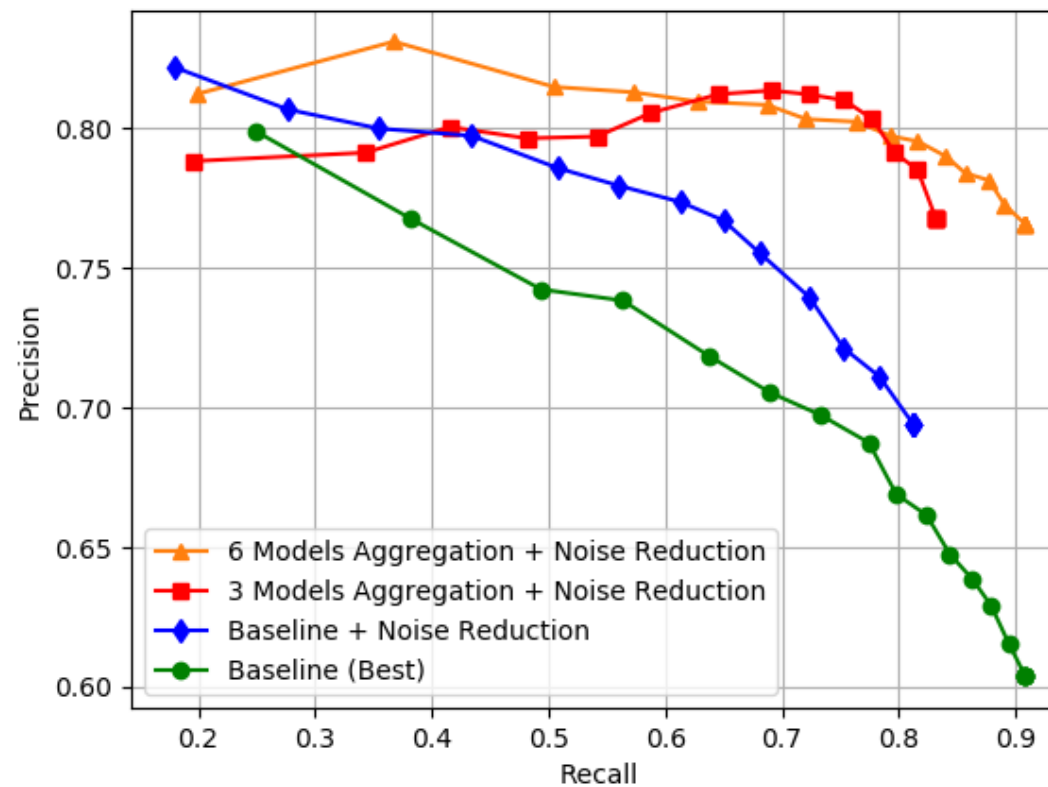


BLACKJACK
38.1%

ROULETTE
21.6%

10

Performance of Tomobox Model for Abusive



The recipe:

- Choose your algorithms w.r.t data
- Do you need a real-time solution?
- Have a good software architecture!
- Flexibility is key
- Feedback loops are your friends
- Get real - evaluate everything

Thank you!

igalg@tomobox.co