

P O L I M I  
D A T A S C I E N T I S T S

# Data Science @ Moviri

Our industrial perspective



The presentation will start at **17:45**



The aperitivo at **19:00**



# What is PMDS?



# Community

- ◎ Build a community of passionate students and professionals about Data Science
- ◎ Exchange ideas and useful tools among members of the community



# Events

- ◎ Connect the community with professionals and corporations whose core business is related to Data Science
- ◎ Connect the community with researchers that operate in the Data Science field



# Projects

- ◎ Take part in workshops organized by companies where professionals will guide you through your first hands on Data Science experience
- ◎ Get the chance to apply your knowledge and skills by working on a real Data Science project



# Data Science resources

- ◎ Open podcast with guests from the Data Science industry and academia
- ◎ Blog posts about ML techniques and tools
- ◎ Monthly newsletter with the most exciting news related to the Data Science field



# Join the community

Connect with us on:

- [polimidatascientists.it](http://polimidatascientists.it)
- [LinkedIn](https://www.linkedin.com/company/polimi-data-scientists/)
- [Instagram](https://www.instagram.com/polimidatascientists/)
- [Telegram](https://t.me/polimidatascientists)



# Stay tuned

This will be the last event before the exam session, but prepare yourself for the event season of the next semester!

## ■ The Moviri Group:

- Corporate Profile

## ■ ContentWise:

- What does and what offers

## ■ Streaming market:

- Fraud detection

## ■ Performance Auto Tuning:

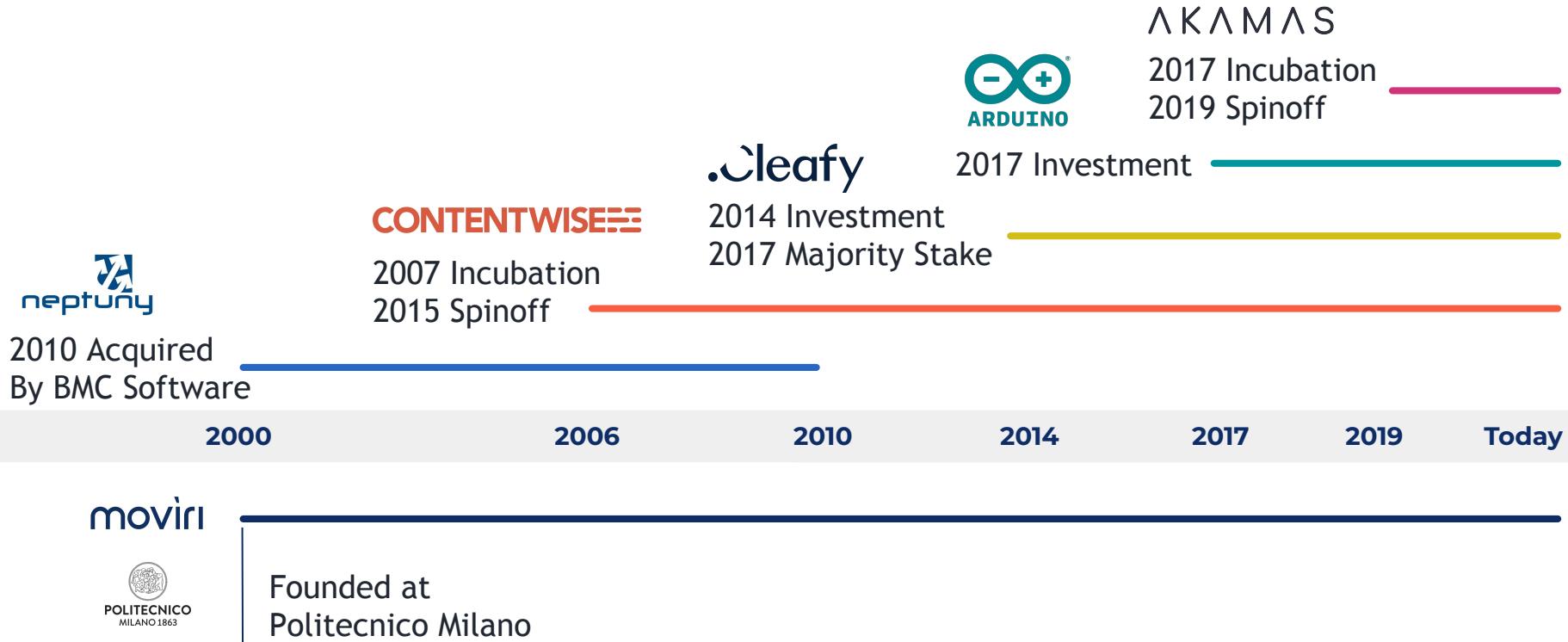
- Bayesian Optimization

POLIMI  
DATA SCIENTISTS

# The moviri Group

Moviri is a multinational technology consulting and software group that uses data, software, and insights to solve substantial business challenges in four strategic fields: **performance engineering, analytics, cybersecurity** and IoT.

# Serialized product innovation



# A Global Footprint

**11**  
Locations

**280+**  
Customers

**30+**  
Countries

**40%**  
International  
Revenues



# ContentWise

## What we do and what we offer

**Filippo Ciceri**  
Tech Lead  
**@ ContentWise**



## Market Problems in Our Domain

1. **Management & Promotion**  
millions of contents each year manually
2. **UX Diversity & User Taste Curation**  
without any evidence
3. **Impact Analysis** based only on high level KPIs
4. **Manual curation** of item features

## Our Mission

We help businesses **engage customers** at scale  
through **AI-powered** digital experience technology



**Adapt** UX and UI  
**to match** a user's tastes and intent

**Enrich**, connect, contextualize and create superior data  
**to improve** the service

**Measure** the effectiveness with AI-assisted technology  
**to drive** business decisions



# Feedback Loop



EDITORS



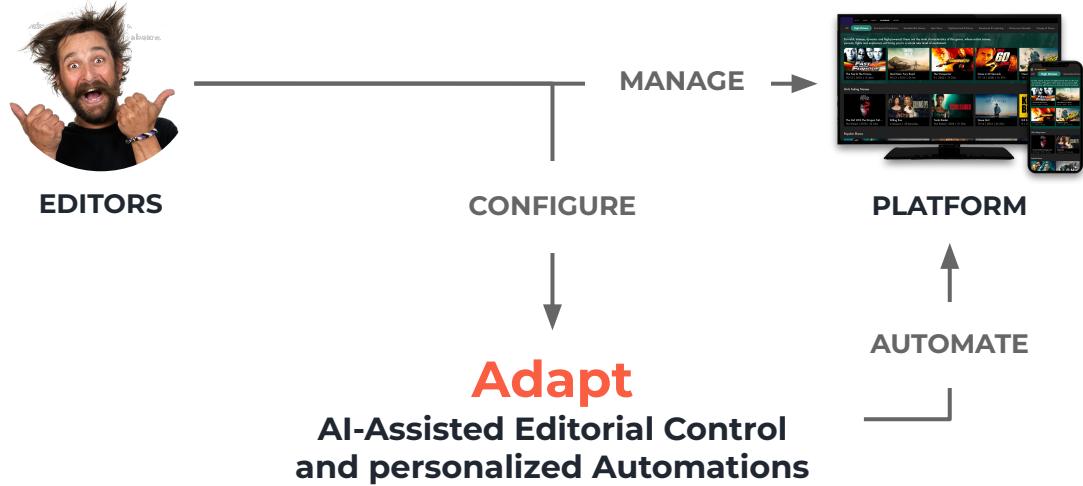
MANAGE →



PLATFORM



# Feedback Loop

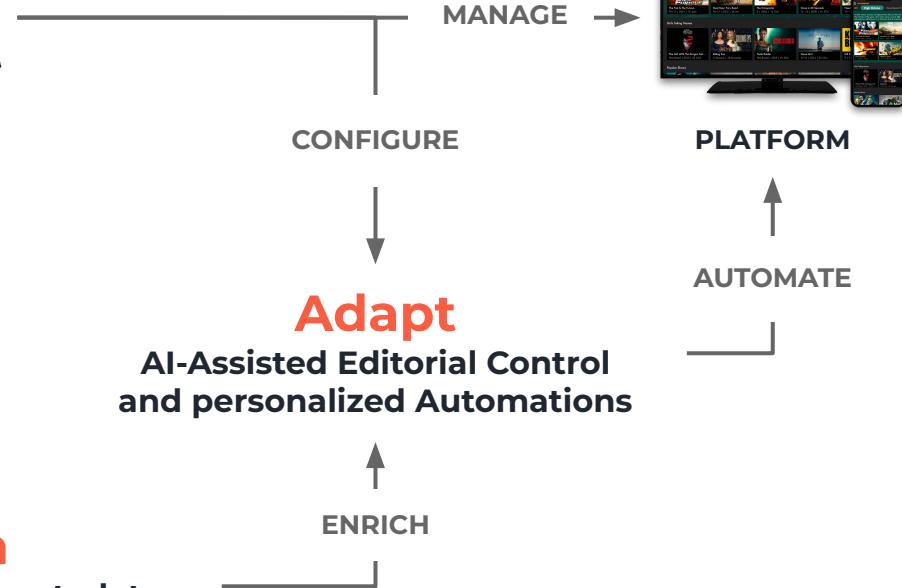




# Feedback Loop



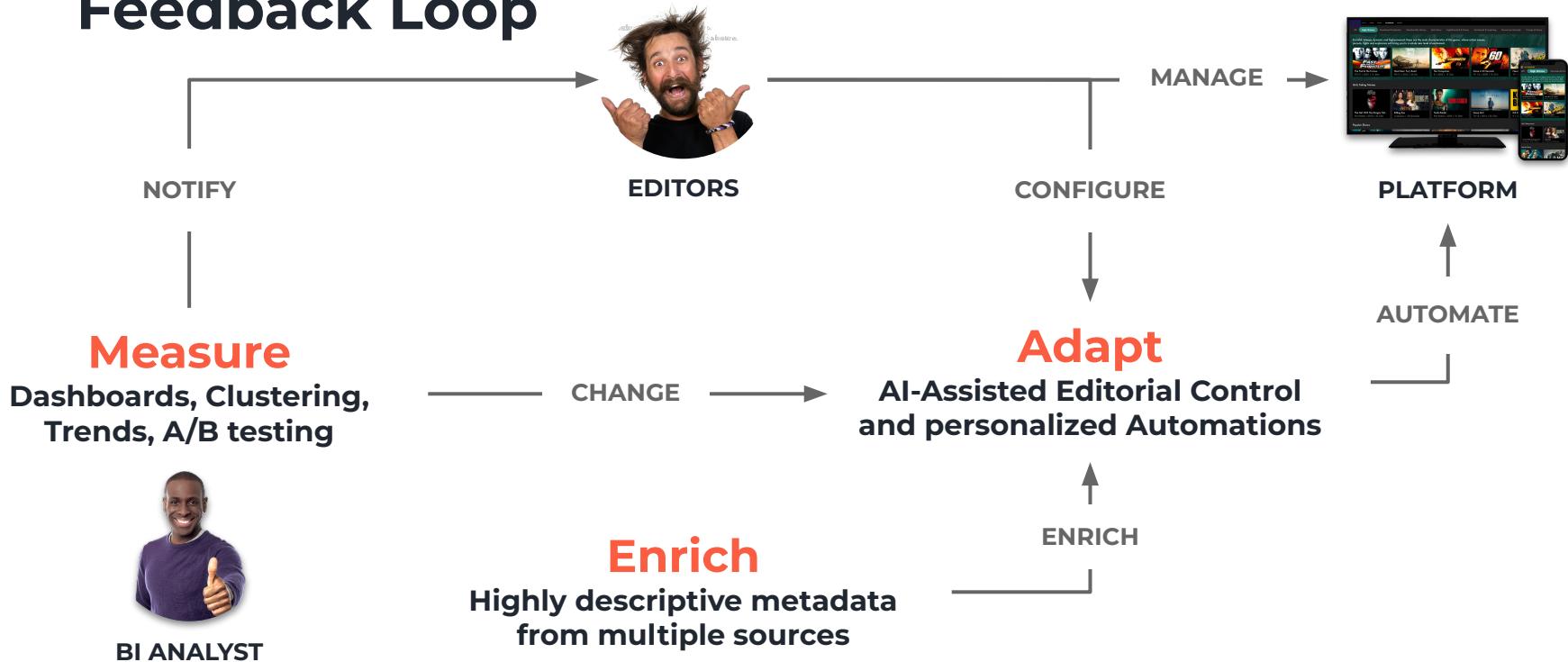
EDITORS



**Enrich**

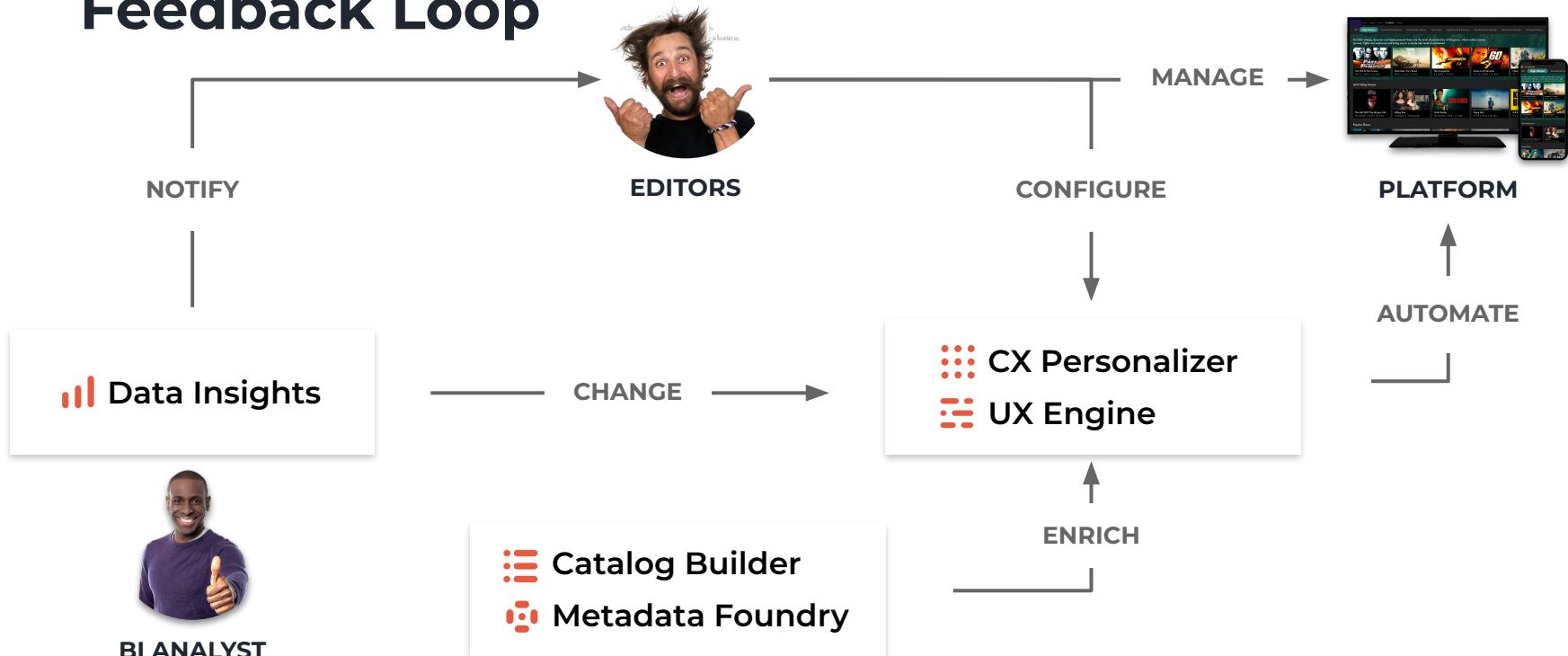
Highly descriptive metadata  
from multiple sources

## Feedback Loop





# Feedback Loop



## Area of Research

### Artwork Generation and Personalization

Generate multiple **artworks** from video

Recommend artworks that best **match** the user profile

### Session Based Personalization

Capture **short term intention** and mood of users

### Enhance Metadata with feature extractions

Extract **features** from videos

### Micro-clustering and Behavioral Analysis

Enable **fine-grained marketing campaigns**

# Who Chose Us

**GEDI**  
GRUPPO EDITORIALE

**EE TV**  
TIMVISION

**RINASCENTE**

**BBC**  
Worldwide

*Telefonica*

**CHILI**  
LIVE YOUR MOVIE

**maxdome**

**AT&T**

**SK broadband**

**multimedia**  
polka

**Canal Digital**

**U** UNITED  
GROUP

**telenor**

**Bell**

**BRUNELLO  
CUCINELLI**

**Rai**

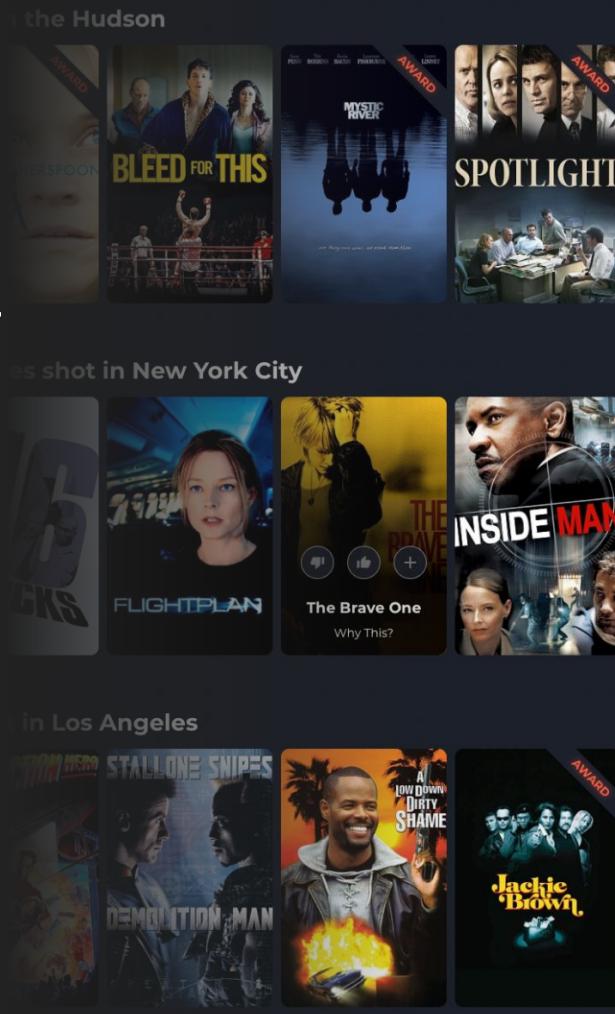
**allente**

**tct**

**PCCW**

**sky**

**& Many More!**



# Analytics Streaming Market Fraud Detection

**Davide Santambrogio**  
Data Scientist  
**@ Moviri**



# Moviri Analytics



## Data Engineering:

- Cloud Architectures
- ETL Pipelines
- Data Ingestion



## Data Science:

- Fraud Detection
- Sentiment Analysis
- Anomaly Detection
- Synthetic Data

# Anti-Fraud Use Case



## Business need

Automated definition of profiles' risk based on **streaming**, purchase and **log-in** behaviour.



## Approach

Building a platform to run analyses on the **risk scores** computed by the Models.



## Issues

Few labeled data of frauds.  
Finding new frauds patterns different from static rules.



## Solution

A **Machine Learning Engine** aiming at optimizing the Risk Measures computation for fraud detection.

# Anti-Fraud Solution - ML approach

Anti-fraud engine based  
on multiple ML techniques

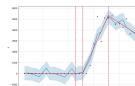
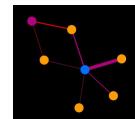
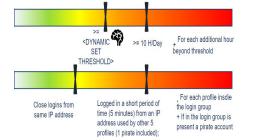
- Streaming data
- Purchase and login data
- Pirates label

## Machine Learning Engine for Fraud Detection

Semi-supervised models

Graph Analysis

Behaviour Clustering



Data Integration & Storage platform

Antifraud Engine

## Risk Score Dashboard for Fraud Analysis



RISK SCORING  
FUNCTION

Data Visualization & Alerting System

moviri

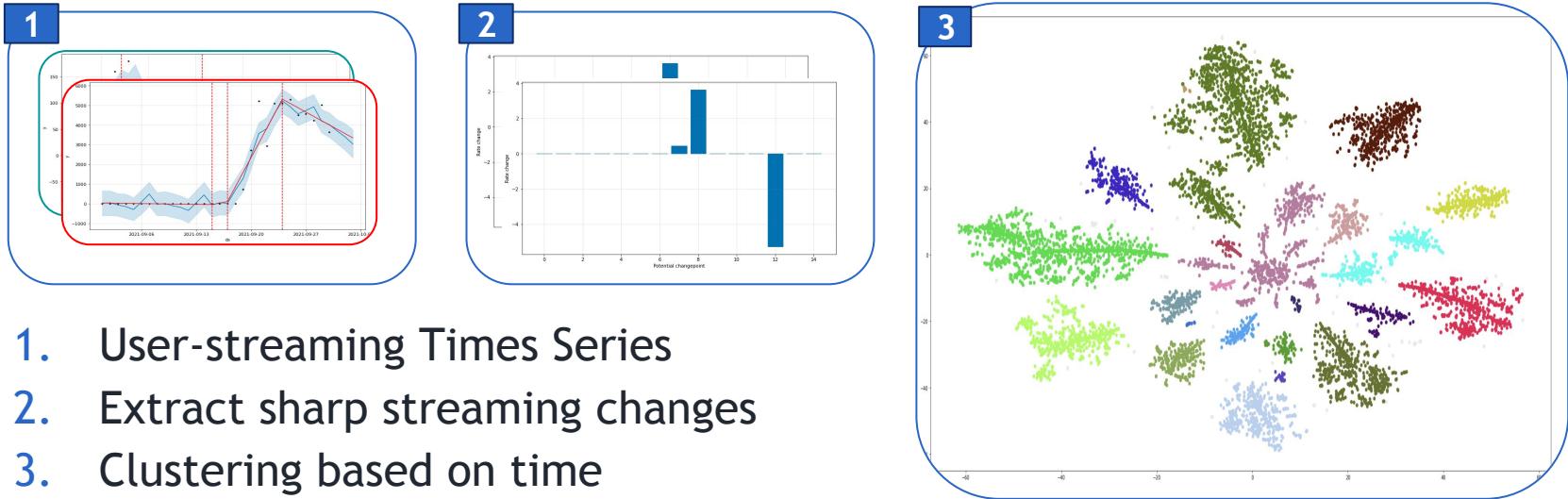
# **How do we find new pirates' patterns?**

**What's the main issue with unsupervised solutions?**

# What we have done so far...

1. Data exploration to define possible patterns
  - a. Enable safe model **explanability** and ease model **validation**
2. Use of **time-series analysis** to derive user embeddings
3. Performing **unsupervised clustering** to group similar users
4. **Validate cluster properties**, e.g. **Domain Knowledge**

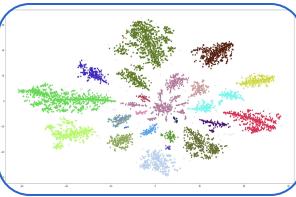
# Behaviour Clustering example



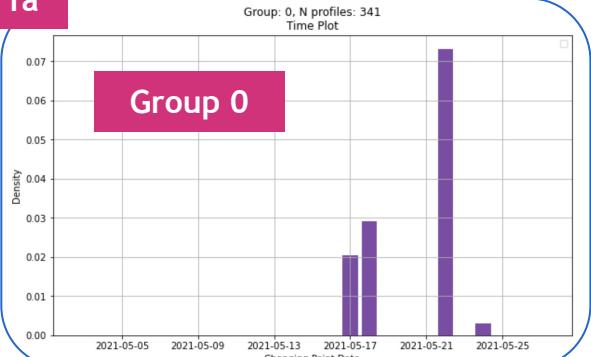
1. User-streaming Times Series
2. Extract sharp streaming changes
3. Clustering based on time and changes magnitude

**Result:** Define groups of similar users, based on their daily time of streaming

# Cluster characterization



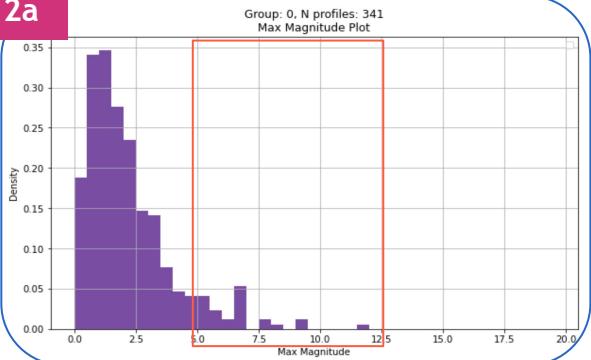
1a



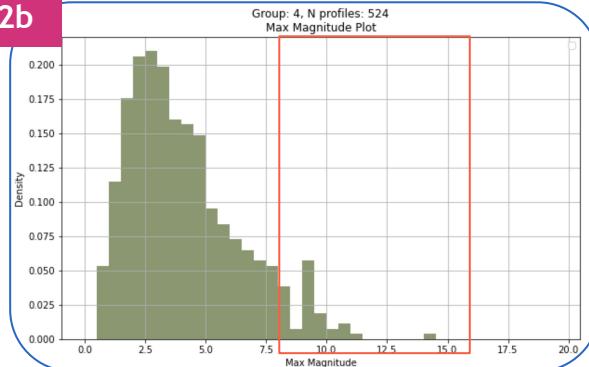
1b



2a



2b



- Clusters with sharp changes on precise days (1a, 1b).

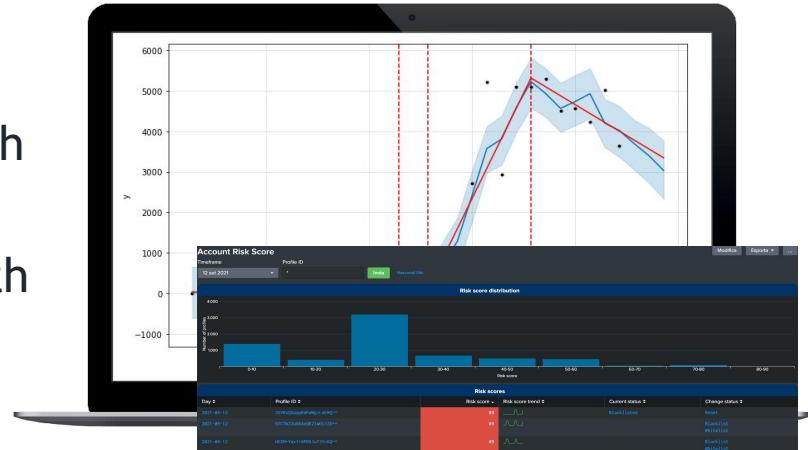
- Consider sharpest changes in streaming (2a, 2b).

# Conclusion

## Fraud detection system based on multiple ML approaches

### Solution

- Labels for semi-supervised models with feedbacks
- Unsupervised behaviour clustering with Domain Knowledge Validation
- Risk score to run Fraud Analyses



**50%**

New fraud profiles identified

**Ranking**

Fraud priority based on risk score

**92%**

Recall on found pirates  
**moviri**

# Akamas

## Performance Auto Tuning

### Bayesian Optimization

**Stefano Cereda**

Senior Data Scientist

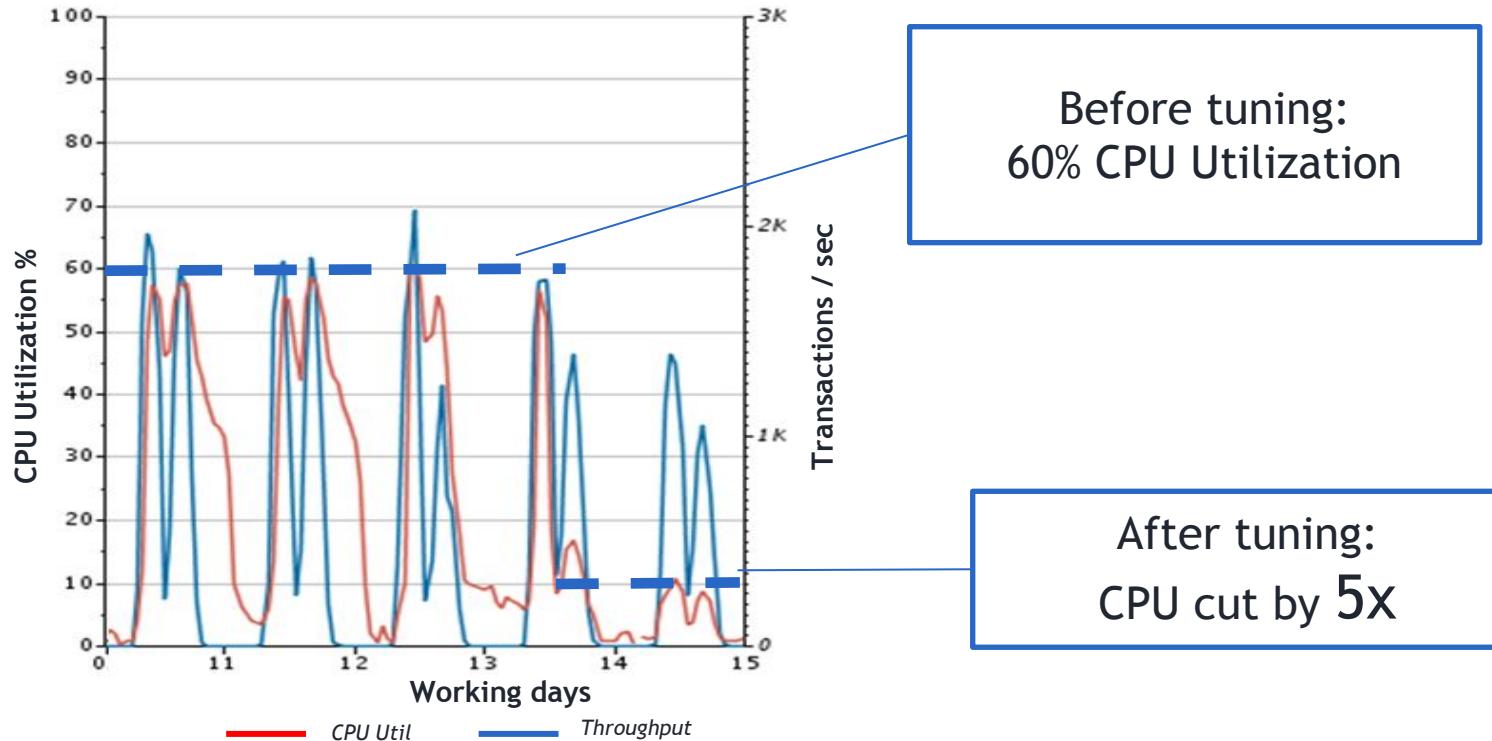
@ Akamas



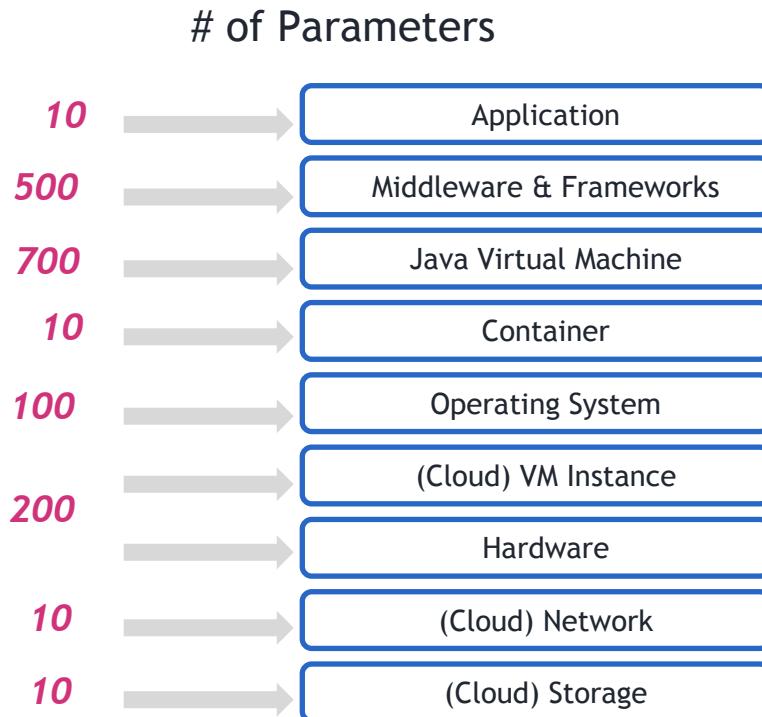
# Akamas

- Configuration of IT systems is becoming more complex
  - Explosion of configuration parameters
  - Frequent releases
  - Microservice architecture
- Negative consequences
  - Overspending
  - Poor performance
  - Low agility
- Akamas provides AI-powered, automated, full-stack and goal-driven configuration autotuning

# Configuration Tuning Results



# Tuning Space is Beyond Human Scale

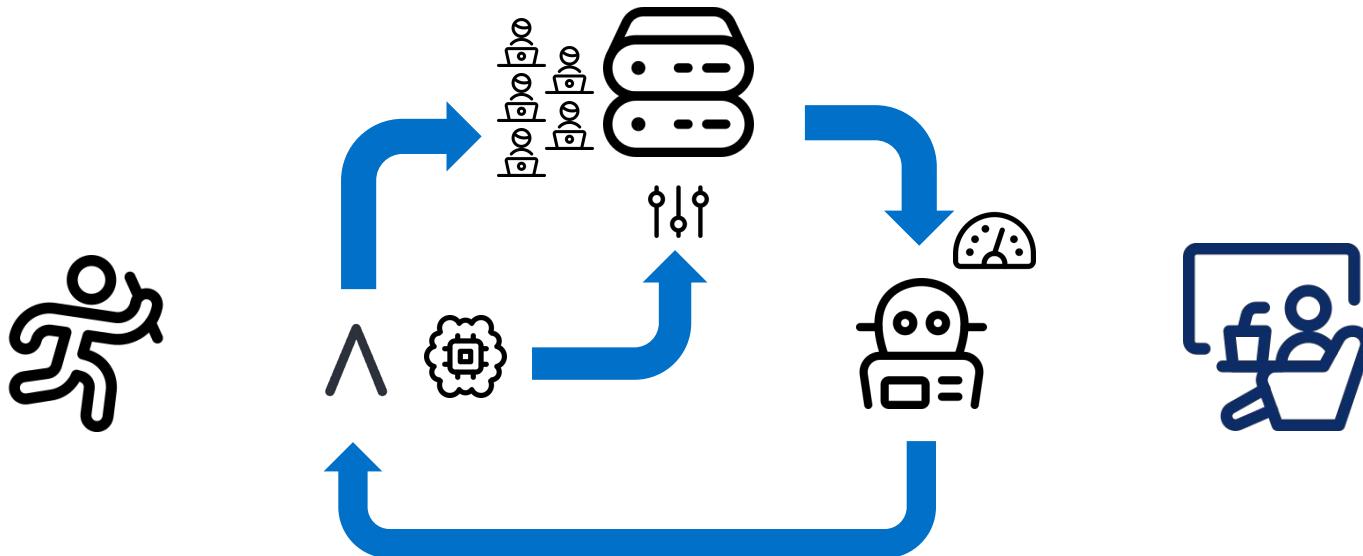


Looking for the  
optimal settings? It's  
easy, just try

$2^{100} =$

121,267,650,600,  
228,229,401,496,  
703,205,376  
configurations...

# Continuous Performance Optimization



# Autotuning Approaches



Model Based  
Queuing Networks  
Petri Networks  
Linear Programming

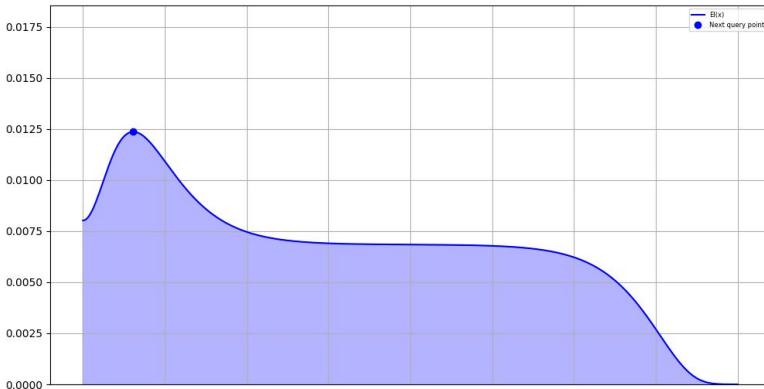
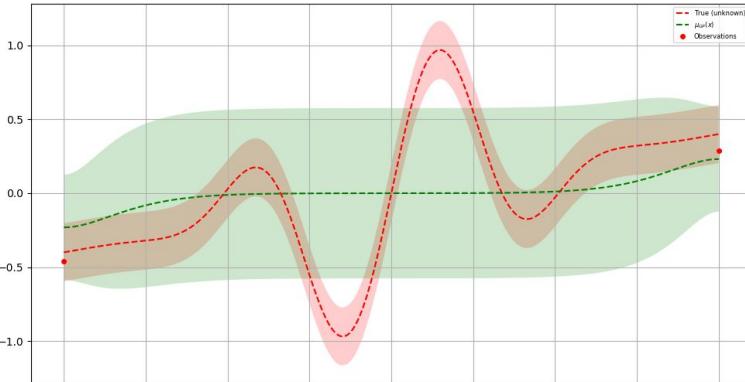


Simulation Based  
Random Forests  
Statistical Machine  
Learning



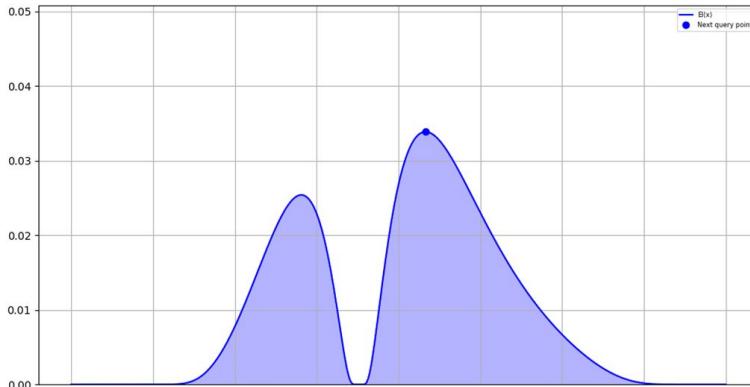
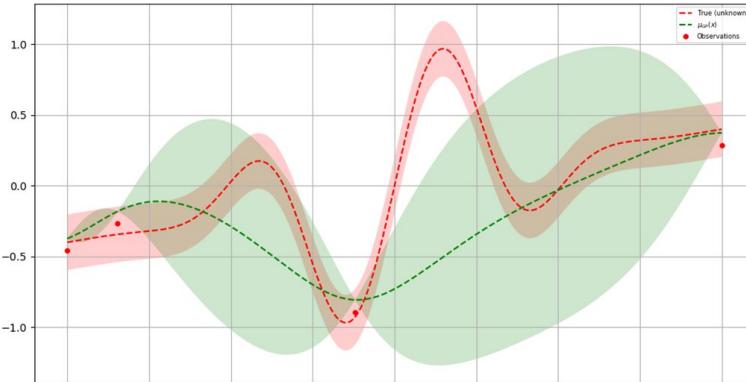
Test Based  
Random Search  
Reinforcement Learning  
Bayesian Optimization

# Bayesian Optimization in a nutshell



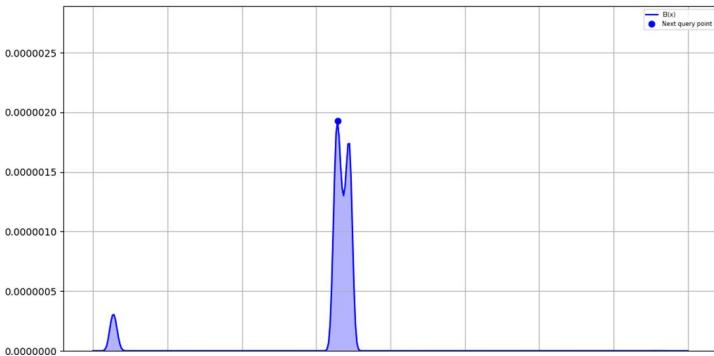
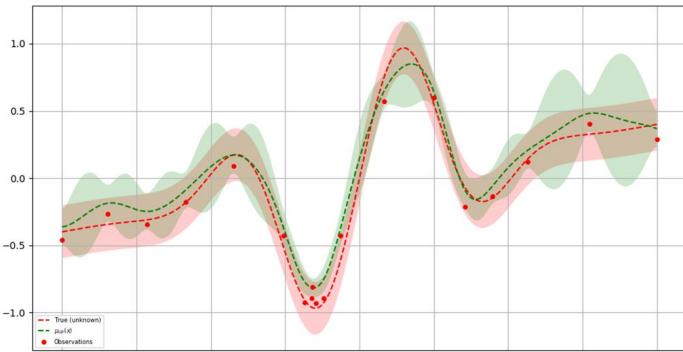
Tested configurations: 2  
Best so far: -0.46

# Bayesian Optimization in a nutshell



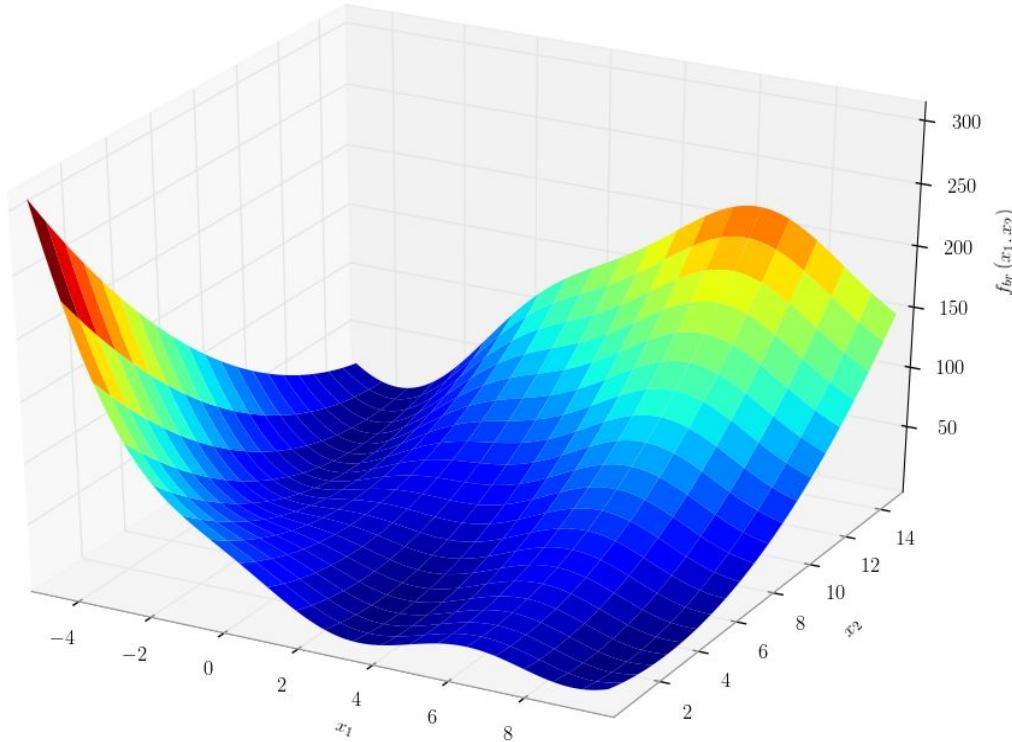
Tested configurations: 4  
Best so far: -0.89

# Bayesian Optimization in a nutshell

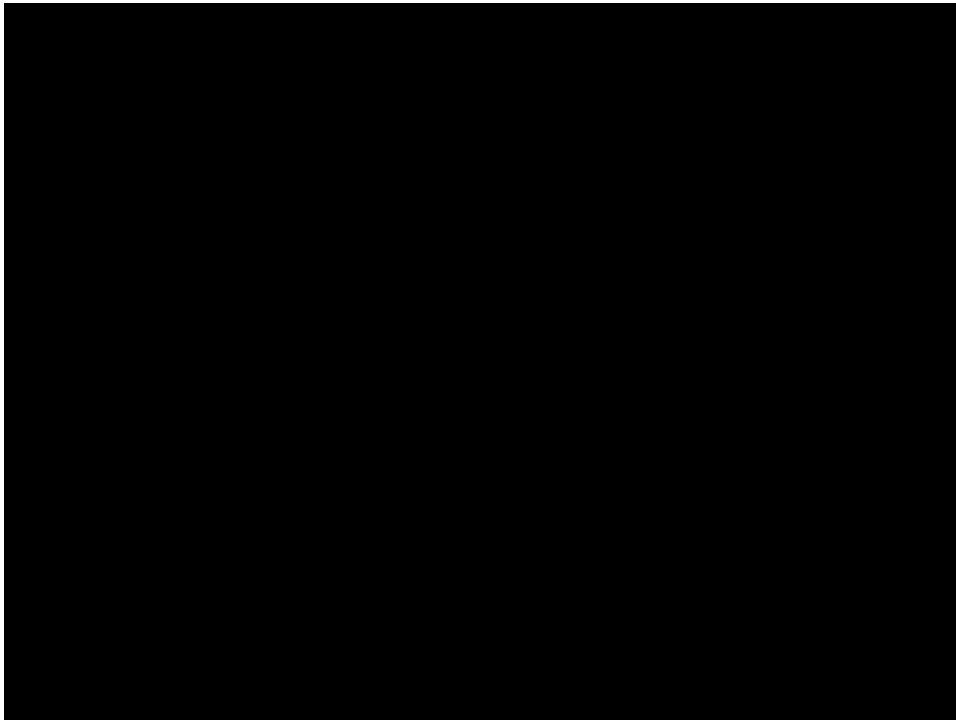


Tested configurations: 19  
Best so far: -0.93

# Bayesian Optimization in Action

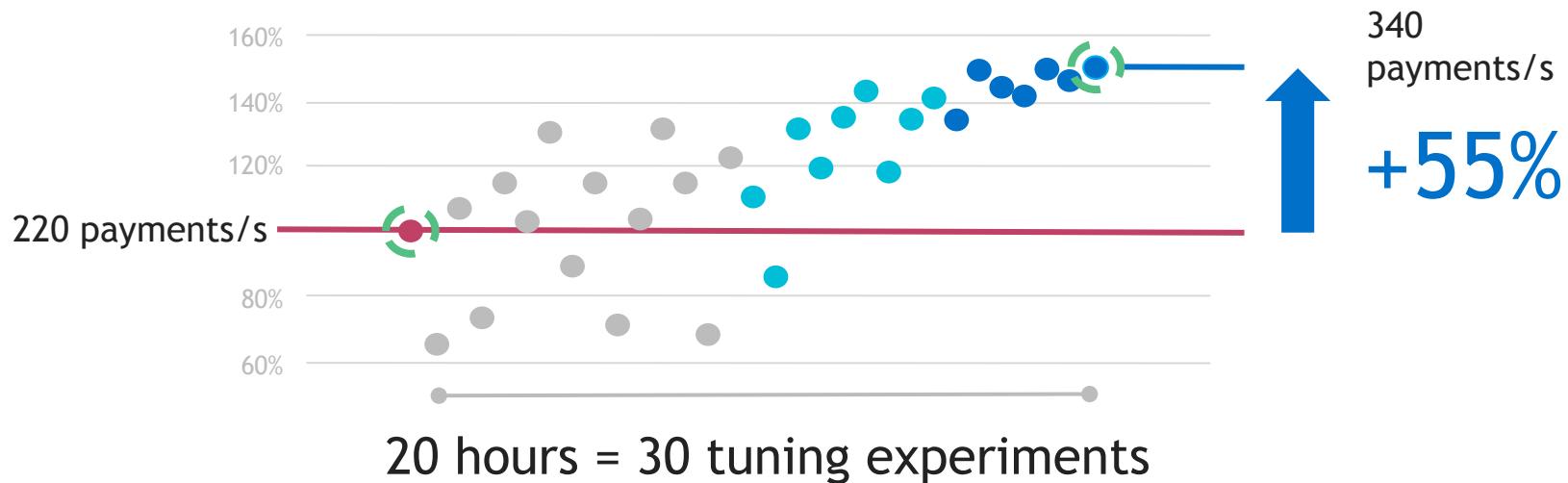


# Bayesian Optimization in Action



# Customer Case: Payment Application Optimization

Goal: maximise payments per second by tuning application, middleware and OS configurations



# Additional Complexity - Thesis

- Performance is a function of both the applied configuration and the **incoming workload**
- Simultaneously tuning interconnected services creates noise
- We usually have constraints on parameters, metrics and exploration
- Some model-derived knowledge might be useful
- Tuning output must be interpretable

# Job Openings

**Junior Performance Engineer** Milan or Padova

**Data Engineer** Milan or Padova

**Cybersecurity Analyst** Milan or Padova

**Junior Frontend Engineer** @ContentWise (thesis opportunity)

Apply on our website: **[careers.moviri.com](https://careers.moviri.com)**

# Let's connect over aperitivo!

HR TEAM



**Annamaria Sciré**  
[annamaria.scire@moviri.com](mailto:annamaria.scire@moviri.com)



**Mara Orianí**  
[mara.oriani@moviri.com](mailto:mara.oriani@moviri.com)



**Maria Giulia Mastrelli**  
[mariagiulia.mastrelli@moviri.com](mailto:mariagiulia.mastrelli@moviri.com)



**Federica Ballancin**  
[federica.ballancin@moviri.com](mailto:federica.ballancin@moviri.com)



### HQ Milan

Via Schiaffino, 11  
20158 Milan

### Padua

Via Sant'Andrea, 21  
35139 Padua

### Boston

211 Congress Street  
Boston, MA 02110

### Los Angeles

12130 Millennium Dr  
Los Angeles, CA 90094

### Singapore

5 Temasek Boulevard,  
Singapore 038985

