# MLOps: are we there yet?

Gianluca Campanella

13th April 2019

# Hello!

My name is **Gianluca** [dʒanˈluːka]

# What I do nowadays

I’m a Data Scientist at

Microsoft

in **AzureCAT**

# What I do nowadays

I also run my own company

Estimand

that provides

**Data Science consulting and training**

# Who's in the room?

# Today we're talking about…

MLOps

# Today we're talking about...

MLOps

($\approx$ a whole bunch of mistakes I made in the last few years)

# Two types of Data Science

## Analysis-focused

- Maths and Statistics
- Business Intelligence

→ Assist human decision-making

## Building-focused

- Machine Learning
- Software Engineering

→ Develop and deploy data-driven products

# Things I’ve helped build recently

- High-frequency trading system for sports betting
- Context-aware, personalised search engine
- Content recommender for a mobile app
- Automated forecasting tool for an e-commerce business

# DevOps

**What?**

Automation practices between software developers and IT

**Why?**

To build, test and release software faster and more reliably

# DevOps

> *At its essence, DevOps is a culture, a movement, a philosophy.*
>
> — *Atlassian*

# MLOps

**What?**

Automation practices between data scientists, data engineers, software developers and IT

**Why?**

To build, test, release and monitor software that embeds ML faster and more reliably

# MLOps

High-risk, high-reward innovation culture

$\downarrow$

Iterate quickly $\longleftrightarrow$ Fail fast

# Why does this matter?

If it's not used in production…

**It never happened!**

# Why does this matter?

If *it is* used in production…

**It better work!**

# Why does this matter?

**As a Data Scientist, MLOps**…

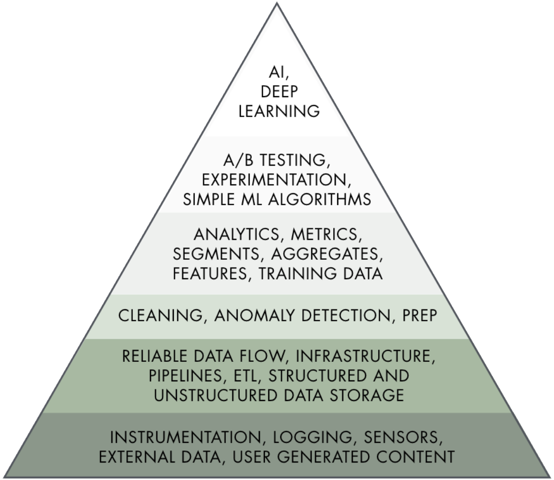
- Is hard but does pay off
- Gives you peace of mind
- Allows you to focus on more interesting tasks

# Why does this matter?

**As a Software Engineer, MLOps**…

- Is something you're probably already doing
- Increases the dependability of ML systems
- Brings you closer to the Data Science team

# Don’t try to run before you can walk




From M. Rogati

# Where are we in the DS workflow?

$$\begin{array}{ccccc} & & \text{Research question} & & \\ & & \downarrow & & \\ \text{Obtain} & \longleftrightarrow & \text{Explore} & \longleftrightarrow & \text{Model} \\ & & \downarrow & & \\ & & \text{Operationalise} & & \end{array}$$

# Where are we in the DS workflow?

Research question

$\downarrow$

Obtain $\longleftrightarrow$ Explore $\longleftrightarrow$ Model

$\downarrow$

Operationalise

# Where are we in the DS workflow?

Research question

$\downarrow$

Obtain $\longleftrightarrow$ Explore $\longleftrightarrow$ Model

$\downarrow$

Operationalise

# Where are we in the DS workflow?

Data sources

$\downarrow$

ETL $\longleftrightarrow$ Model development $\longleftrightarrow$ Training

$\downarrow$

Deployment

# Where are we in the DS workflow?

$$
\begin{array}{ccccc}
 & & \text{Data sources} & & \\
 & & \downarrow & & \\
\text{ETL} & \longleftrightarrow & \text{Model development} & \longleftrightarrow & \text{Training} \\
 & & \downarrow & & \\
 & & \text{Deployment} & &
\end{array}
$$

# ETL

**What?**

- Extract, transform, load
- Data Science alchemy
- Heavily informed by exploratory data analysis (EDA)

# ETL

**Things to keep in mind**

- Distributional assumptions
- Transformations
- External data sources

# Model development

**How?**

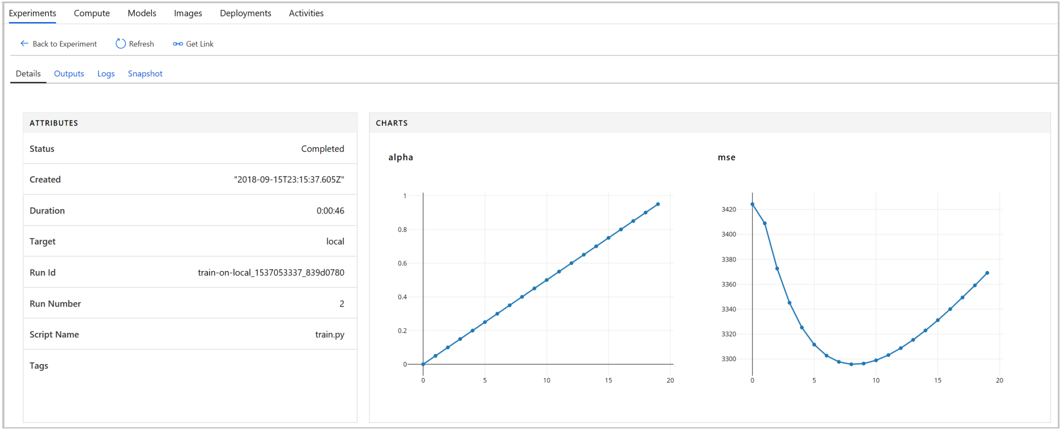
- Hyperparameter tuning
- Automated ML

# Model development

**Things to keep in mind**

- Choice of programming language
- Versioning
- Performance tracking

# Model development



# *Interlude*

**Online vs offline metrics**

- Business metrics $\rightarrow$ online metrics
- Offline metrics $\approx$ online metrics
- Experiment early and often

# Interlude

**Feedback loops**

- Models become self-fulfilling prophecies
- Biased data collection
- Don’t just exploit, explore

# Training vs scoring

**Training**

- Historical data $\rightarrow$ model
- Scheduled offline (batch) jobs

# Training vs scoring

**Training**

- Historical data $\rightarrow$ model
- Scheduled offline (batch) jobs

**Scoring**

- Model + new data $\rightarrow$ predictions
- Online or offline (batch)

# Deployment

**How?**

- Offline (batch) scoring
- Queues
- RPC (e.g. REST endpoint)
- In-process

# Deployment

## Things to keep in mind

- Throughput and latency requirements
- Impedance mismatch between training and scoring*
- Access control and security
- Other moving parts (e.g. databases)

* I'm looking at you, Apache Spark

# Online deployment

**How?**

- Docker
- Kubernetes
- CI/CD pipeline

# Online deployment

**Things to keep in mind**

- Splitting traffic
- A/B testing
- Incremental roll-out

Data sources

$\downarrow$

ETL $\longleftrightarrow$ Model development $\longleftrightarrow$ Training

$\downarrow$

Deployment

So… we're done?

# Not quite!

We still need to automate ETL and training

# Retraining

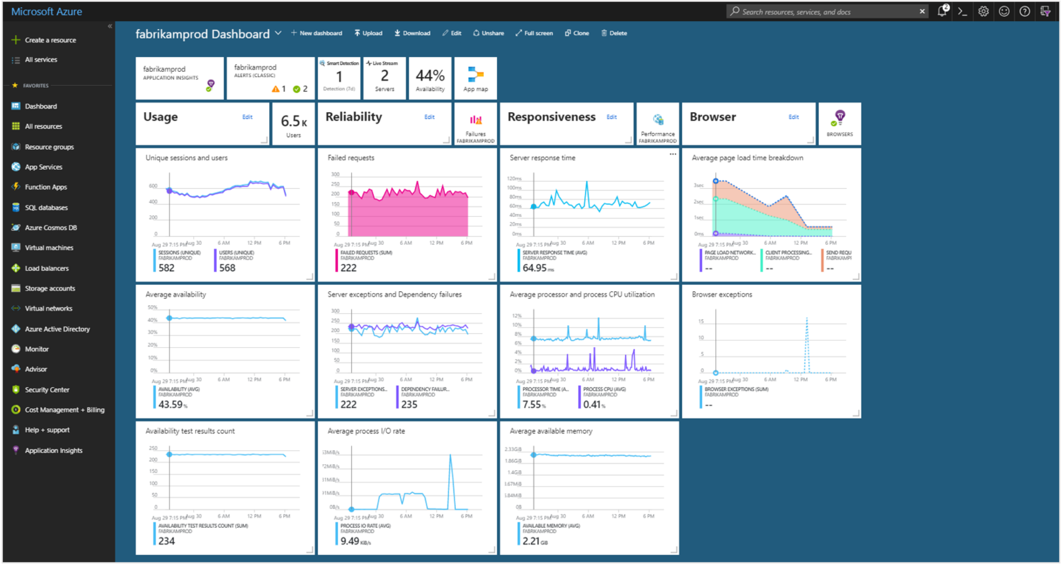
**Things to keep in mind**

- Distributional assumptions
- Data drift
- Performance tracking
- Golden set

# Monitoring

- Logging
  - Distributional assumptions
  - Data drift
  - Statistical performance
  - Serving performance
- Anomaly detection and alerting
- Fallback mechanisms

# Monitoring


Microsoft Azure
fabrikamprod Dashboard
Create a resource
All services
Dashboard
All resources
Resource groups
App Services
Function Apps
SQL databases
Azure Cosmos DB
Virtual machines
Load balancers
Storage accounts
Virtual networks
Azure Active Directory
Monitor
Advisor
Security Center
Cost Management + Billing
Help + support
Application Insights
Usage
Reliability
Responsiveness
Browser
Unique sessions and users
Failed requests
Server response time
Average page load time breakdown
Average availability
Server exceptions and Dependency failures
Average processor and process CPU utilization
Browser exceptions
Availability test results count
Average process I/O rate
Average available memory

Now we’re really done!

But then...

# Do it again!

- Versioning
- Roll-back mechanisms
- Experimentation
- Bandits

# Recap

$$
\begin{array}{ccccc}
 & & \text{Data sources} & & \\
 & & \downarrow & & \\
\text{ETL} & \longleftrightarrow & \text{Model development} & \longleftrightarrow & \text{(Re)training} \\
 & & \updownarrow & & \\
\text{Experimentation} & \longleftrightarrow & \text{Deployment} & \longleftrightarrow & \text{Monitoring}
\end{array}
$$

# Recap

**As a Data Scientist…**

- Familiarise yourself with the tools
- Try moving some of your workloads away from your laptop
- Understand where Engineering is coming from

# Recap

**As a Software Engineer**…

- Ramp up on containers and orchestration
- Check out the different cloud offerings
- Help your fellow Data Scientists

# Thank you!

If you want to keep in touch…

@ gianluca@campanella.org   gcampanella   gcampanella