CERN IT Machine Learning Infrastructure Workshop

Welcome and Introduction

Sofia Vallecorsa, Ricardo Rocha

https://indico.cern.ch/event/1253881/

Motivation

Machine Learning (ML) at CERN since LEP years

- Limited abstraction
- Preliminary feature engineering
- Mostly for classification & regression

Deep Learning (DL)

- Increasingly abstract representations
- Complex computational graphs requiring larger infrastructure
- A large range of applications

ML / DL has different requirements than typical HEP software

Enabling Deep Learning

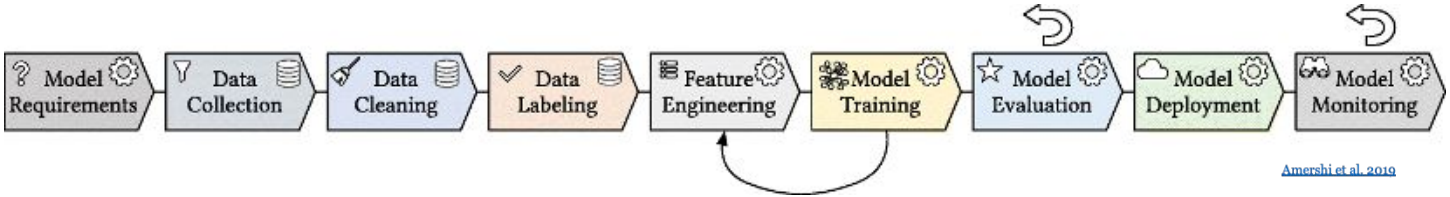
Productive R&D goes hand-in-hand with **smart infrastructure design**.

Developing AI systems is **different** than typical software engineering. In particular:

- **data** discovery, management, and versioning are **more complex**;
- **modular design** is **not trivial** since AI components exhibit **complex entanglement**.

However **deployment** & **integration** into “regular” software and services is critical

- Already 10 years of **“AI lifecycles”** - inspired by data mining standards



Initiative

Improve the Machine Learning offering in CERN IT

Understand what is there and what is missing

Follow up with improvements to services and end user experience

Proposed Activities and Workshops

1. **Today**: Survey the ML infrastructure offering and usage in CERN IT
2. TBA: Survey and document use cases and requirements CERN wide
3. TBA: Survey and document state of the art technologies and infrastructure

Today’s Goals

Survey CERN IT ML infrastructure and services

Collect the department use cases for ML

Prepare for the Microsoft workshop happening April 4th. Joint research on:

- Heterogeneous Architectures
- Data Management and Storage
- **AI Models and Platforms**
- *Digital Twins*

Today’s Agenda

https://indico.cern.ch/event/1253881/

- Service Sessions: 20 minutes + 5 minutes Q&A
- Application Sessions: 7 minutes
- Discussion and Next Steps

Next Steps

To be reviewed together at the end of this workshop…

Organize future workshops

- Use cases and requirements CERN wide
- State of the art technologies and infrastructure

Propose ML initiatives with all business engagement units

https://www.xkcd.com/303/

Data Scientist

www.elbruno.com

Thanks everyone for participating!