



# My experience as Data Scientist at BIX

**Digital Innovation Lab @ Boehringer Ingelheim**

Elena Bruna  
*CERN Career Networking event*  
*Nov 15, 2021*

# My career as physicist

- Master (2003) and PhD Thesis (2007) in ALICE @ LHC
- Post-doc in STAR @ RHIC (2008-2010)
- Researcher in ALICE (2011-2018)



## Main activities:

- Simulations for ALICE ITS and EMCal
- Open charm analyses in heavy-ion collisions
- Jets and B-tagged jets

# What is BI X

BI X is the Digital Lab of Boehringer Ingelheim

## Goals

1. Develop innovative digital products for better healthcare
2. Drive Boehringer Ingelheim's digital transformation

## Examples

Software to improve pharma production

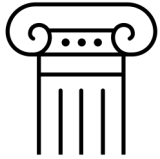
Applications to help human/animal lives

Medical device software

# What is Data Science (@Bi X)?

Data science is a combination of ...

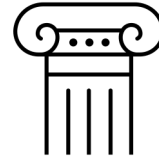
**Machine learning &  
statistics & optimization**



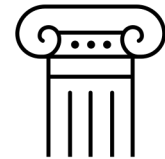
**Computer science &  
software engineering**



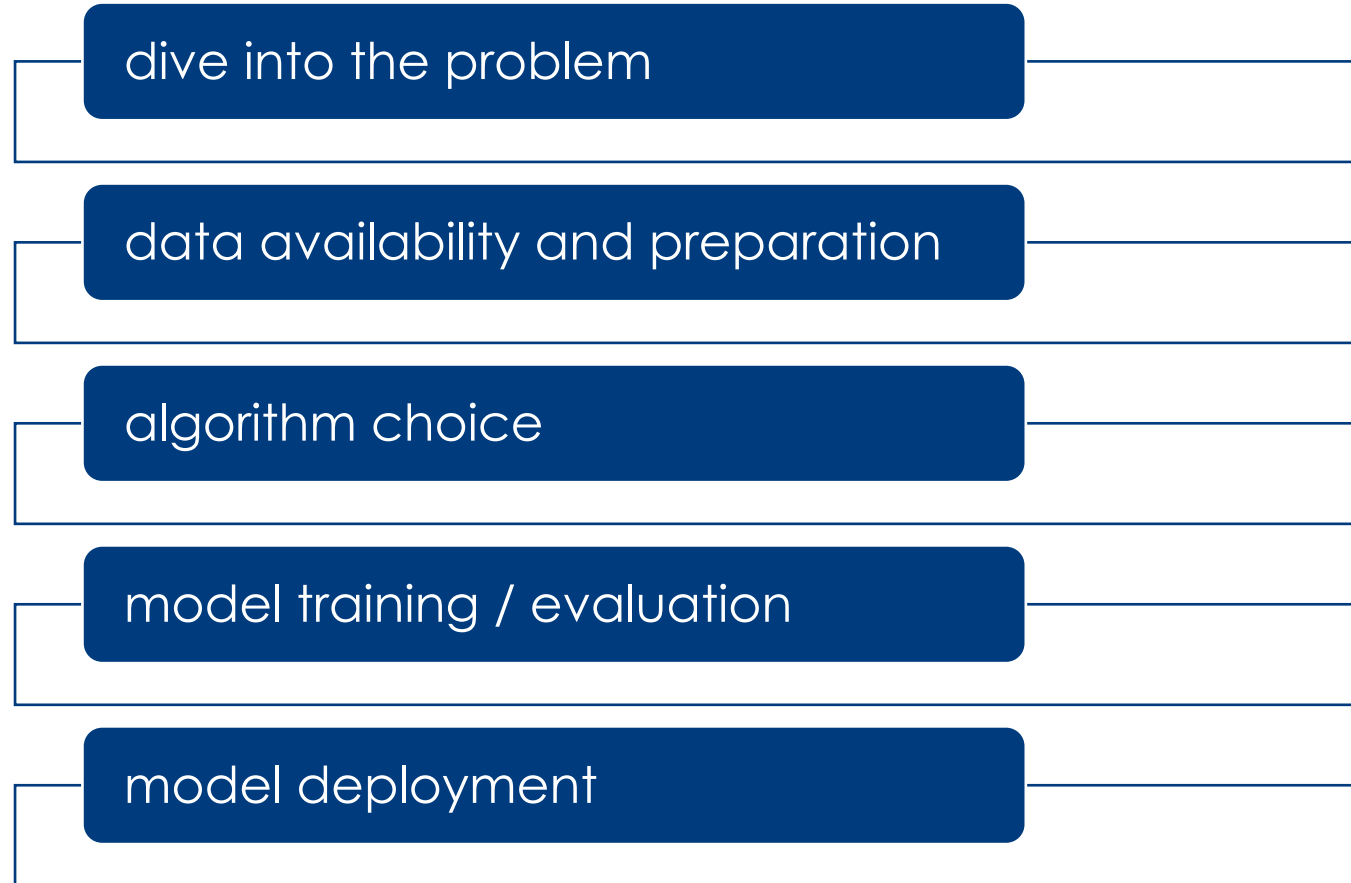
**Business  
understanding**



**Communication &  
visualization**



# The Data Scientist's journey



# Physicist vs Data Scientist

- Based on my experience -

## High-energy physics

- data structure usually designed in advance
- large data availability

## other fields (pharma)

- value of data not always known a priori:
- data not often available, or little, or messy
  - exploratory data analysis crucial
  - often a plan to collect data is needed



DATA

# Physicist vs Data Scientist

- Based on my experience -

High-energy physics

other fields (pharma)

ALGO  
DEVEL.

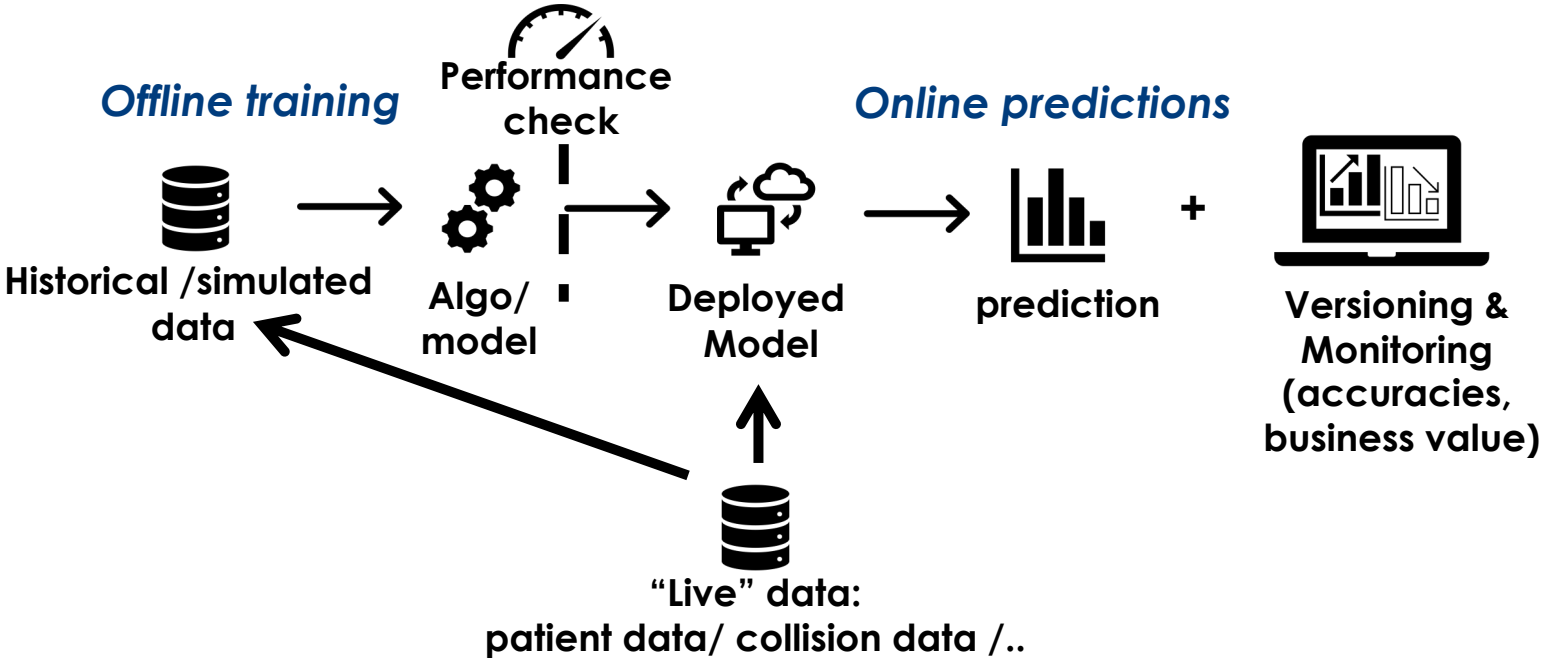
Data can be simulated (MC truth) to develop model/algo

- Choice of algo depends on available data
- Not always time for deep dive and/or best choice
- Model should be easy to explain (e.g. need of transparency in medical device software)

# Physicist vs Data Scientist

- Based on my experience -

High-energy physics ~ other fields (pharma)



MODEL DEPL.



# Physicist vs Data Scientist

- Based on my experience -  
(on a more subjective level)

On one side...

- Missing research atmosphere, discuss results in large communities
- Personal 'attachment' to physics
- Less freedom to work on a given topic

on the other side...

- Apply analytical skills to projects and see them being used in real world
- Connection to science (Pharma/chemistry)
- Expand skills by working in different projects
- Smaller teams, more focused, quick support, fast iterations

## Skills (often asked)

- python
- databases (e.g. SQL)
- data analysis on distributed resources (e.g. Spark)
- analytical/critical approach
- presentation/visualization of results (e.g. Dashboard)

## Suggestions for the CV

- two pages max and well organized
- emphasise skills & tools like :
  - big data analysis
  - ML techniques (with examples)
  - Monte Carlo simulations
  - experience in different projects/analyses
  - work in international teams

## Skills (often asked)

- python
- databases (e.g. SQL)
- data analysis on distributed resources (e.g. Spark)
- analytical/critical approach
- presentation/visualization of results (e.g. Dashboard)

## Suggestions for the CV

- two pages max and well organized
- emphasise skills & tools like :
  - big data analysis
  - ML techniques (with some examples)
  - Monte Carlo simulations
  - experience in different projects/analyses
  - work in international teams

Get **familiar with tools** you are not using on a daily basis

**Find parallels with your work in physics** (programming language, databases,...)

**Underline that you know the concept**, already used a similar tool and can quickly catch up

# Some suggestions for the DS interview

**Prepare** on some work where you used some **DS-specific techniques** (classification of rare signal, clustering hits in detector, ...)

Discuss the **steps** (data preparation, algorithm testing, deployment, ...), **bottlenecks** and **results: communication is crucial**

- Be quantitative but explain in a high-level way

Keep in mind: often in business environment it is not clear what is the “**value**” of a given project in academy. **Find parallels**, like:

- Online/offline data reconstruction → there will be N users/physicists using the reconstructed data for their research

Be **open and confident** that with your experience you can always learn new tools

Thank you!