

Master in Data Science - Lesson 1

Introduction

Mi presento



Formazione

- Laurea in Ingegneria Informatica (Italia)
 - Università di Modena e Reggio Emilia (triennale)
 - Politecnico di Milano (magistrale)
- Dottorato in Machine Learning / Bioinformatica
 - Université libre de Bruxelles (Bruxelles, Belgio)

Esperienza lavorativa

- Centro di Genetica Umana dell'ospedale Erasme (Belgio)
 - Bioinformatico & Ricercatore
- Aspire Financial Technologies Pte. Ltd (Singapore - Remoto)
 - Startup fintech company
 - Data Scientist & Engineering Manager
- Data Scientist Freelancer (Remoto)

La figura del Data Scientist

Al servizio del Business Team

- Il nostro ultimo fine è di essere di supporto per Business a vendere prodotti e servizi dei clienti



Charlie You (ML Engineered)

@CharlieYouAI

...

The customer doesn't care that you're using the latest, greatest, SotA transformer model.

The customer only cares that it does what you said it will do, that it adds value to their life.



Deni Moka ⚡ @dmokafa · Jan 2

The measure of a successful software project is not about:

- number of lines of code
- number of interconnected services
- number of features
- number of applied bleeding and cutting edge technologies

It is about how happy the customer is.

1:25 AM · Jan 4, 2021 · Twitter for iPad

1 Retweet 11 Likes

La figura del Data Scientist

Al servizio del Business Team

- Il nostro ultimo fine è di essere di supporto per Business a vendere prodotti e servizi dei clienti



- Python e ML sono strumenti importanti ma non gli unici
- Altri linguaggi di programmazione (R, SQL, Bash), competenze di sviluppo di codice, visualizzazione dei dati, conoscenza del dominio, etc.. sono altrettanto rilevanti



Travis Gerke
@travisgerke



Spotify Wrapped is proof that sometimes count(x, sort=TRUE) is all the data science you need


♡ 3,105 11:21 PM - Dec 6, 2020



💬 333 people are talking about this



Good vs Great data scientist



Jon Krohn
 @JonKrohnLearns

...

What separates a good data scientist from a great one?

8:29 PM · Apr 15, 2021 · Twitter Web App

<https://twitter.com/JonKrohnLearns/status/1382672535091871748>



Mason @Mace_Grim · Apr 15

Agree with this. It's looking at INDIVIDUAL data points where I learn the most about the mechanics of the data. For some reason, that felt unnatural and "slow" at first, but I really find a lot of value in it.

1 2 7



Pawel Mikler @pawelmikler · Apr 16

Replying to @JonKrohnLearns
 I'd say a unique combination of tech skills, maths know-how, storytelling, and intuition. The best #DataScientists I know are great listeners who focus on problem, provide solutions in quick iterations (the don't give a shit about being wrong) and are ultimate communicators.

3 3 43



It was Algebra All Along @kylewadegrove · Apr 16

Replying to @JonKrohnLearns
 A good data scientist doesn't introduce complexity and a great one knows when to.


7



Russell Pollari @russ_poll · Apr 15

Replying to @JonKrohnLearns
 - MVP mindset. Finding the simplest solution
 - Good communication
 - Good software engineering skills

4 3 69



Darvish Lee Shadravan @DShadravan · Apr 15

Replying to @JonKrohnLearns
 Soft skillz. In addition to being a good data nerd are they also talented at speaking, writing, and understanding the business.

1 1 23



Brett Tully @brett_tully · Apr 16

Replying to @JonKrohnLearns
 Most replies here are true of any team member. For DS specifically I would say creativity, intuition, and understanding what these two phrases mean:
 - you can't sow a field turning it over in your mind
 - a week of hard work can save you an afternoon of careful thought

1 1

Data science skill & Software engineering skills



Shreya Shankar
@sh_reya

...

I'm tired of hearing people blame data scientists for broken production ML pipelines because they don't have "good software engineering practices." We can have a more productive conversation. 1/7

https://twitter.com/sh_reya/status/1383117201394126850



Shreya Shankar @sh_reya · 23h
Replying to @sh_reya

...

First, no one (data scientists included) sets out to write irreproducible code. Right now, writing reproducible ML and DS pipelines requires large amounts of iteration on different ideas and can come at the expense of finding a viable modeling solution. 2/7

2

1

54



Shreya Shankar @sh_reya · 23h

...

If data science OKRs are centered around high-performing models, data scientists rightfully should primarily be aiming for good performance. Jupyter notebooks are here to stay, since no other tool (IMO) supports such rapid DS idea testing and iteration. 3/7

3

2

57



Shreya Shankar @sh_reya · 23h

...

Second, data scientists are not software engineers. If we fault data scientists for not understanding software engineering, shouldn't we fault software engineers for not understanding data science? Unclear if broken ML pipeline blame should sit only on DS shoulders. 4/7

2

5

72



Shreya Shankar @sh_reya · 23h

...

Third, we engineers have literally taken courses on code readability, data structures, algorithms, etc, but many data scientists have never taken a computer science course. They may have taught themselves programming as a means to getting experiment results. 5/7

2

4

57



Shreya Shankar @sh_reya · 23h

...

We can have a little more empathy for our DS colleagues whenever they name their notebook (that they didn't intend for anyone else to see) Untitled12.ipynb instead of expecting them to name variables or decompose code perfectly. 6/7

5

2

42



Shreya Shankar @sh_reya · 23h

...

I'm not saying that data scientists shouldn't have good software engineering practices. I'm saying we shouldn't expect it, especially when the tools we've built for them require them to think about good engineering practices rather than their job (analysis, experiments). 7/7

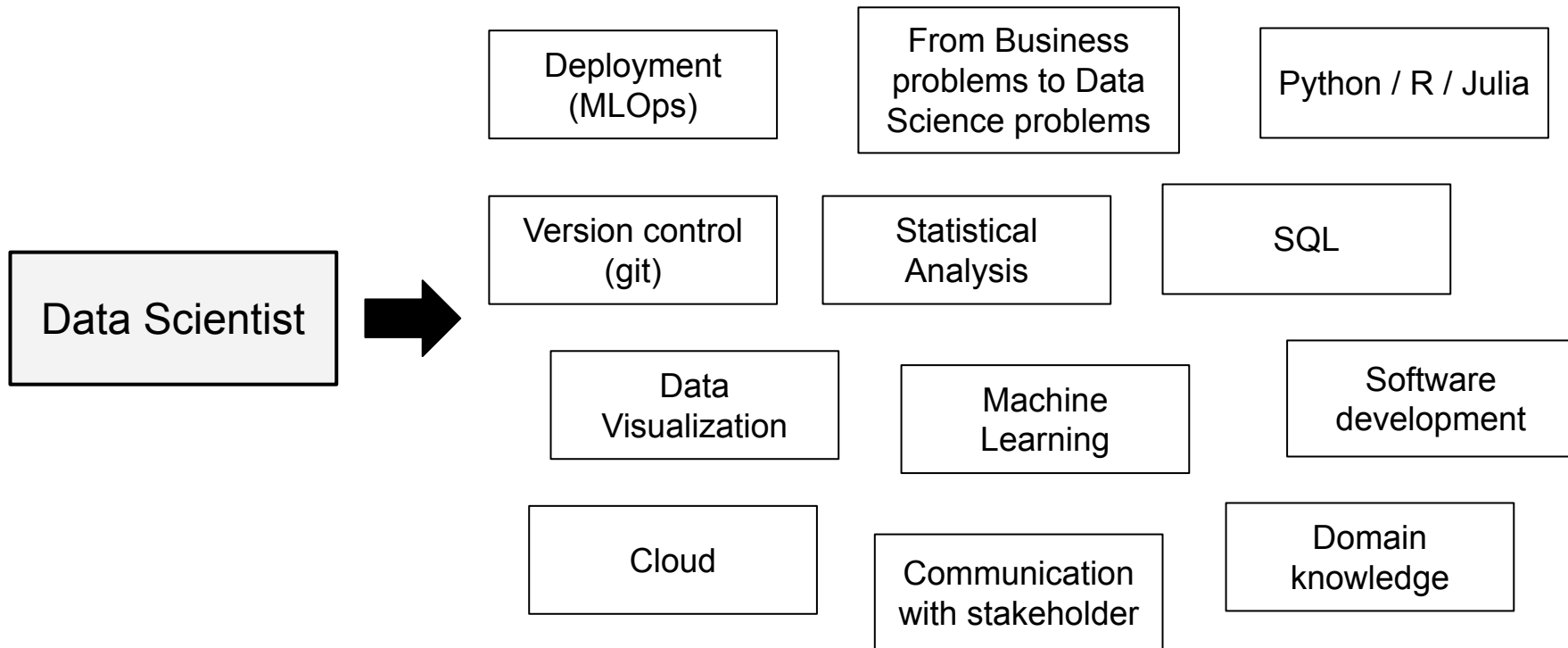
5

5

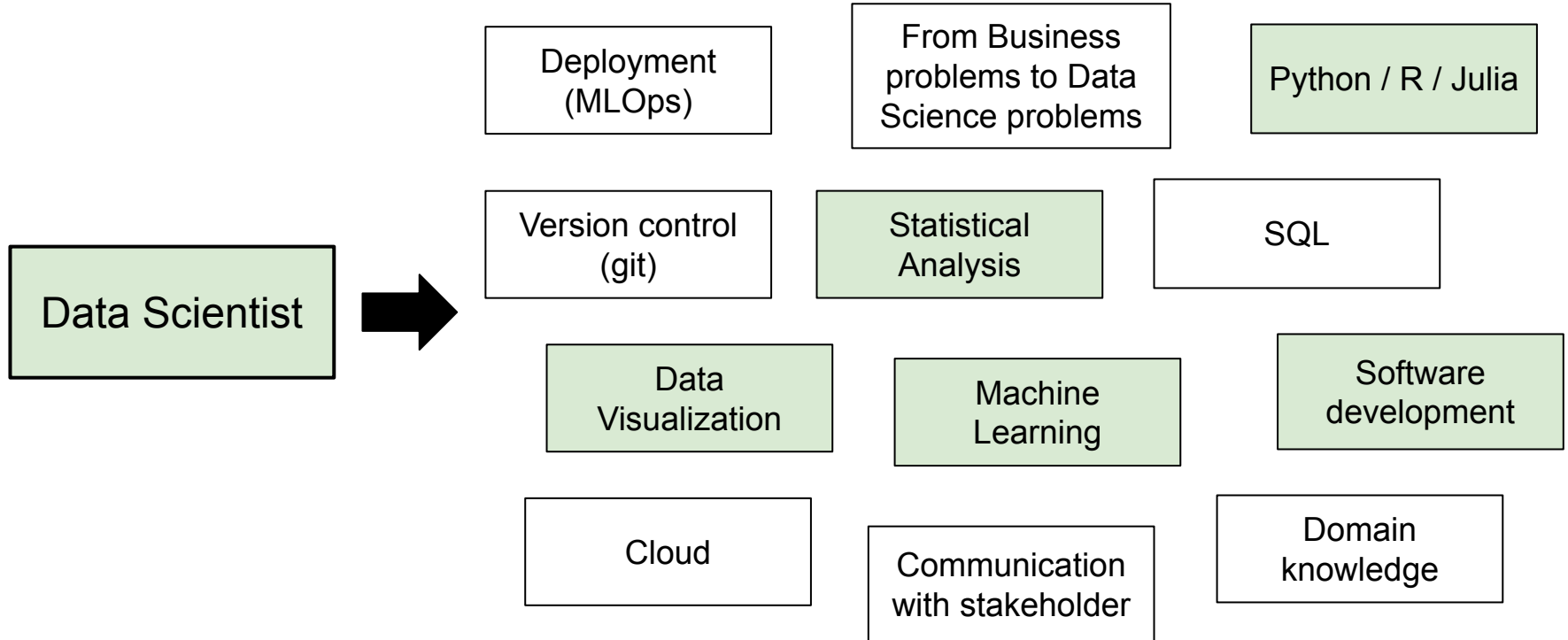
67



La figura del Data Scientist



La figura del Data Scientist



Buono Studio!

Contatti:

claudio.reggiani@neural.academy

agata.manco@neuralacademy.it

carlotta.reggioli@neuralacademy.it