Why do we need Data Scientists?

Antonietta Mira
Co-Founder and Co-Director
InterDisciplinary Institute of Data Science
Università della Svizzera italiana
What is Data Science?

BIG DATA

INFORMATION

INSIGHT

ACTIONABLE KNOWLEDGE
What is big data?

• **BIG** = E-NORMI = EX (out of) NORMA (norm) = EXTRA-ORDINARY

• COMPLEX DATA

• SMART DATA

• **DATA** = DATUS = GIVEN = Immediate, straightforward

• DATA SCIENTISTS *mediate* big data
<table>
<thead>
<tr>
<th>LATITUDE</th>
<th>LONGITUDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>46.94809</td>
<td>7.44744</td>
</tr>
<tr>
<td>46.95334</td>
<td>7.44870</td>
</tr>
<tr>
<td>46.94416</td>
<td>7.44860</td>
</tr>
</tbody>
</table>
WHO EARN WHAT SALARY
NATIONAL AVERAGE SALARIES (US)

STATISTICIAN
$ 75,069

DATA SCIENTIST
$ 118,709

$62,379
$65,991
$67,672

$95,936
$100,118

$116,725
$118,709
Where does big data come from?

**Digitalization**
- multiple **types**: text, image, audio, video, GPS
- multiple **sources**: social, mobile, web, drones, wearables
- multiple **collection techniques**: real time, metadata

**Computation**
- High performance computers
- Quantum computers

**Interconnectivity**
- Human – machine
- Machine – machine
Human – Machine interactions

THE INFINITELY GREAT

STARS

THE TELESCOPE

THE INFINITELY SMALL

CELLS

THE MICROSCOPE

SOCIETY

NATURE

THE MACROSCOPE

THE INFINITELY COMPLEX

Social Science revolution
Data Science revolution

• Education
• Medicine: personalized and precision medicine
• Economics: e-commerce
• Finance: fintech
• Agriculture: drones crop inspection
• Industry 4.0
• ...
Data Science value chain

- Interpret – knowledge domain experts
- Manage – clean, consolidate, store, retrieve
- Visualize and integrate – standardize definitions, harmonize
- Analyze – statistical models, machine learning extract information
- Learn – from information to knowledge
- Act – support decisions and create value
Innovative scientific methodology

Complex models

Efficient algorithms

Powerful super computers

Uncertainty quantification

Pose the right questions

Scientific innovation
A data scientist is someone better at statistics than any software engineer and better at software engineering than any statistician.
Data science educational programs

- **US & CA**
  - 130 Master’s
  - 15 Bachelor’s
  - 11 Doctorates

- **UK and Ireland**
  - 40 Master’s
  - 4 Bachelor’s
  - 1 Doctorate

- **Asia & Oceania**
  - 5 Master’s
  - 1 Bachelor’s
  - 1 Doctorate

- **Europe**
  - 37 Master’s
  - 5 Bachelor’s
  - 2 Doctorate
Data Science: academia & industry

- DS creates a new paradigm to scientific discovery
- DS offers a new way of working within companies and organizations
- DS is creating a new culture in which researchers with different backgrounds join forces to extract knowledge from data
- DS is creating a new culture in which leaders and IT join competences to realize value from data

- Leonardo-style Scientist
- Disciplinary Scientist
- Multidisciplinary Scientist
Big Data leads to Big Science

H2020 Initiative on Big Data
“Large Scale Pilot actions in sectors benefitting from best data-driven innovation such as health, energy, environment, earth observation, geospatial, transport, manufacturing, finance and media.”

SNSF National Research Programme on Big Data:
"The program will address the technical questions raised by Big Data (infrastructures, machine learning, security), study the societal challenges Involved (social acceptance, regulatory aspects and economic impacts) and facilitated the development of new applications.”
Big Data leads to Big Money

We are witnessing a new industrial revolution driven by digital data, computation and automation

US healthcare: $300 billion

EU public sector: $250 billion

US retail: 60%  \(\uparrow\) net margin

Manufacturing: 50%  \(\downarrow\) product development costs

End users: $600 billion from location data

EU Commission communication, Towards a thriving data-drive economy. 2014
Big Data can cause Big Errors/Risks

Commercial interests dominate platforms that generate and traffic Big Data

Legal, confidentiality, and ethical issues

Narrows our choices - False creativity – Herding effect

Inequality because of asymmetry in knowledge management and analysis

Differential privacy

Transparent data process – access protocols

Still need for vision and human insight
EU commission: “Data driven innovation: The capacity of business and public sector bodies to make use of information from improved data analytics to develop services and goods that facilitate every day life of individuals and organizations”

A recent report from McKinsey Global Institute estimates that big data analytics could increase annual GDP in retail and manufacturing in US by up to $325 billion by 2020. By 2018, US will experience a shortage of 190,000 skilled data scientists, and 1.5 million managers and analysts who can handle big data.