



Data Science challenges in 2020: new application domains

13/10/2020 – Rladies

Giulia Brandetti
Senior Data Scientist



Who I am

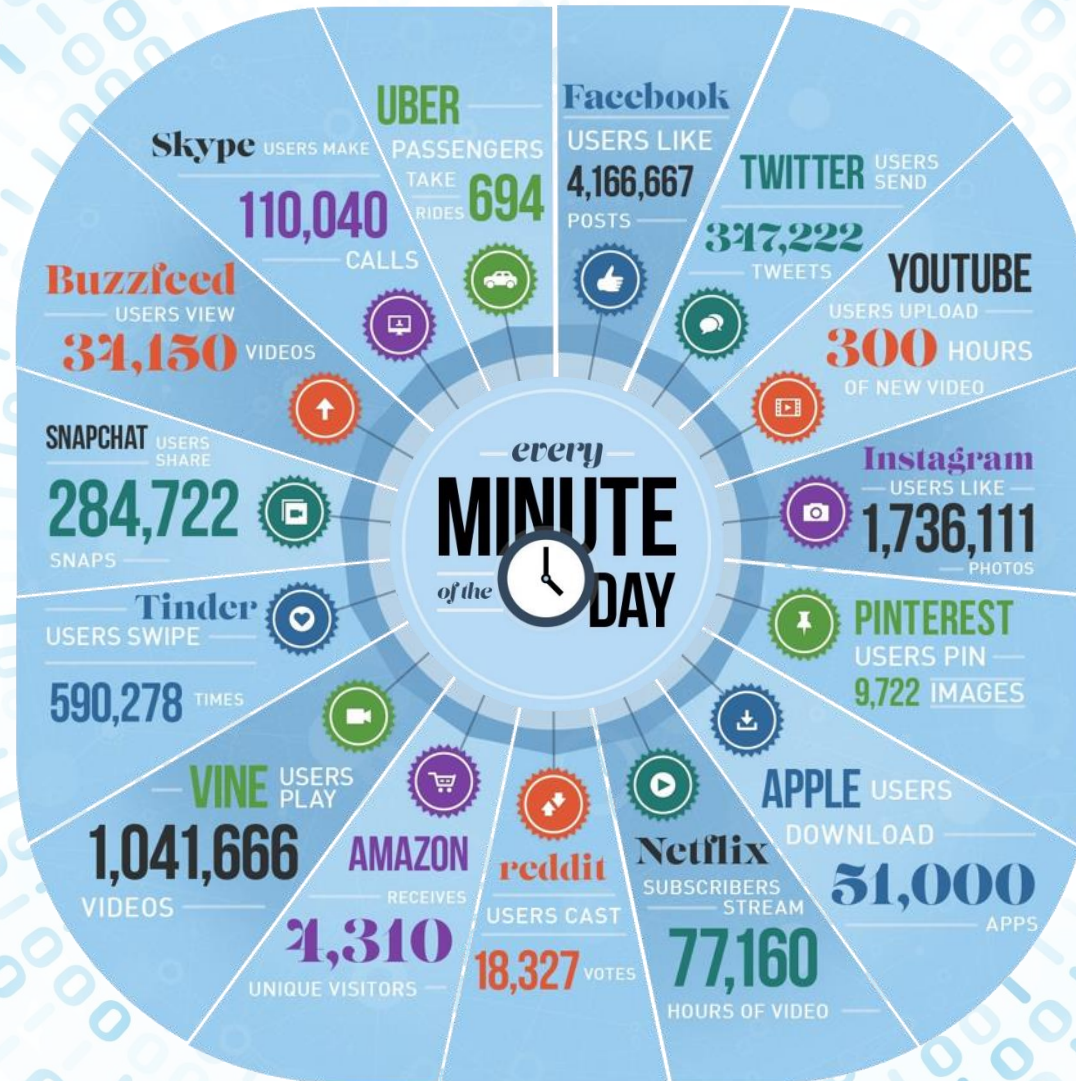


```
{name: 'GIULIA BRANDETTI',  
  role: 'Senior Data Scientist @ Enel',  
  education: 'Master in Statistics',  
  mail: 'giulia.brandetti@enel.com'}
```

"It is a capital mistake to theorize before one has data".

– Sherlock Holmes

Do you know how many things happen in 1 minute?



Data is the new oil

A data driven approach is a key pillar for a digital strategy



What it means being data driven?



Using data to **drive business processes and decisions**



Using **best in class architectures** to provide data in a real time way



Spreading **data culture** at each level of the organization



Fostering **data quality** to have more reliable data



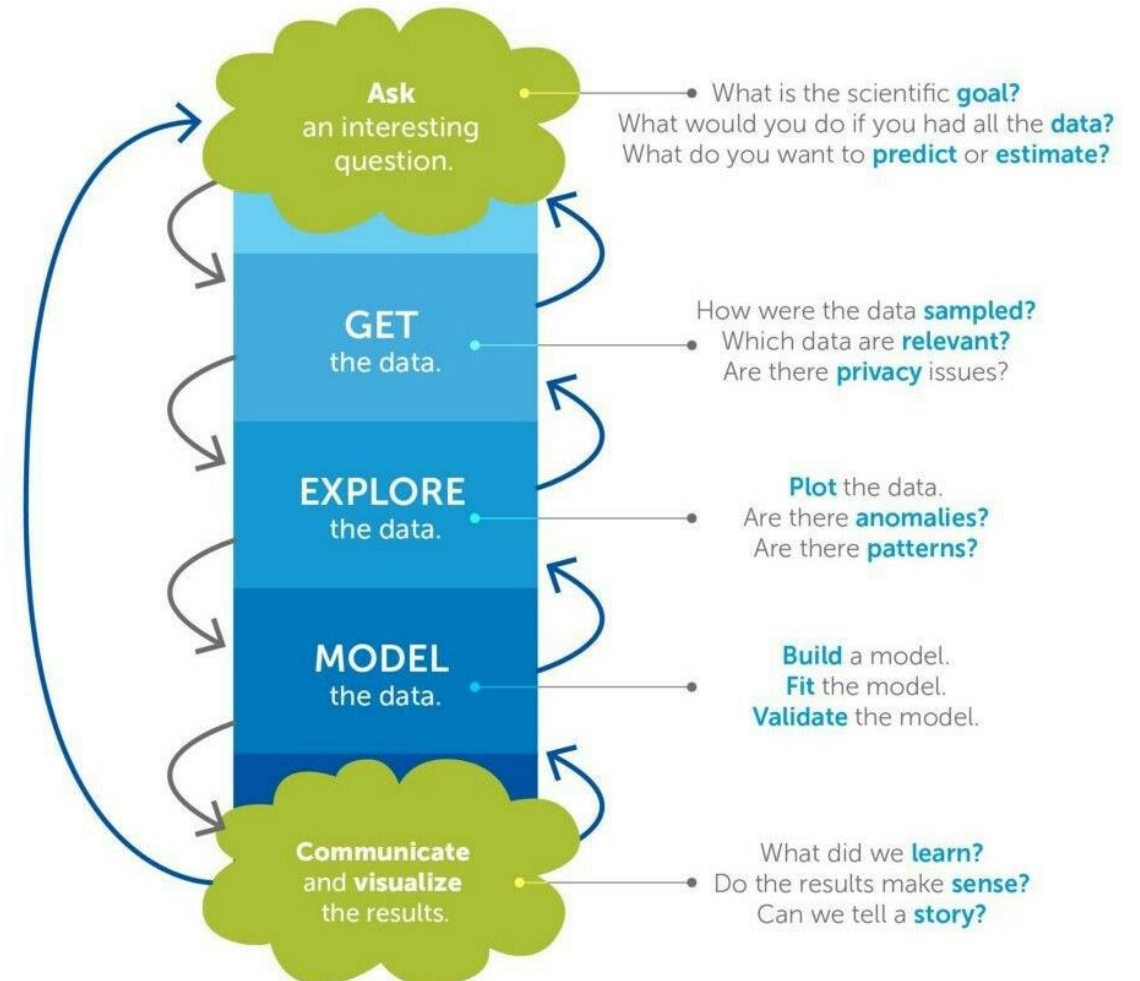
Having a strong **data governance** with clear roles and responsibilities

What is data science



- discovering what we don't know from data
- obtaining predictive, actionable **insight** from data
- creating **Data Products** that have business impact now
- **communicating** relevant business stories from data
- building **confidence** in decisions that drive business value

The Data Science Process



«Machine learning. This is the next transformation... the programming paradigm is changing. Instead of programming a computer, you teach a computer to learn something and it does what you want.»

Eric Schmidt

Executive Chairman of the Board
Google



Data Science applications

We can find data science applications in every business



- Recommendation systems
- Consumers profiling
- Feedback analysis



E-commerce

- Anomaly detection
- Predictive maintenance
- Monitoring systems



Manufacturing

- Customer value
- Fraud detection
- Credit risk



Banking

Healthcare



- Medical Image analysis
- Survival analysis
- Bioinformatics

Transport



- Self Driving Cars
- Cars monitoring systems
- Enhancement of driving ex.

Finance

- Customer segmentation
- Algorithmic Trading
- Strategic decision making

DATA SCIENCE APPLICATIONS

People Analytics: a data science challenge for people management



How Google Is Using People Analytics To Reinvent HR*

Shifting to data-based people management



"All people decisions at Google are based on data and analytics."

Continuous innovation cannot occur until a firm makes a **strategic shift toward a focus on great people management**.

Almost every current HR function operates under 20th century principles of past practices, efficiency, risk avoidance, legal compliance, and hunch-based people management decisions. If you want serial innovation, you will need to **reinvent traditional HR** and the processes that drive innovation.

No one in finance, supply chain, marketing, etc. would ever propose a solution in their area without a plethora of charts, graphs, and data to support it, but HR is known to all too frequently rely instead on trust and relationships. **People costs often approach 60% of corporate variable costs**, so it makes sense to manage such a large cost item analytically.

* Sources: <https://www.ere media.com/tlnt/how-google-is-using-people-analytics-to-completely-reinvent-hr/>, February, 2013
<https://www.entrepreneur.com/article/284550>, November, 2016

Google's answers to 4 key problems



What makes a successful leader?

Periodic one-on-one coaching which included expressing interest in the employee



When and why employee will leave the company?

A retention algorithm allows management to act before it's too late



How to predict upcoming people management problems?

Developing predictive models and "what if" analysis



Which candidates had the highest probability of succeeding after they are hired?

An effective predictive hiring algorithm

Google's people analytics team doesn't dictate; it convinces with data

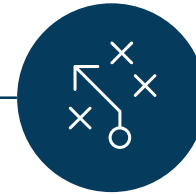
People Analytics @ Enel



Which are the most common causes of the employee unsatisfaction?

HR-BP interview

Text clustering



Sentiment classification

Labelling and visualization

DATA



53k interviews



5+ languages



1 year



142 HR-BPs



22k Italian
texts analyzed

* Removing **punctuation**, **numbers** and **stop words** (300 words, including "Enel", ...)

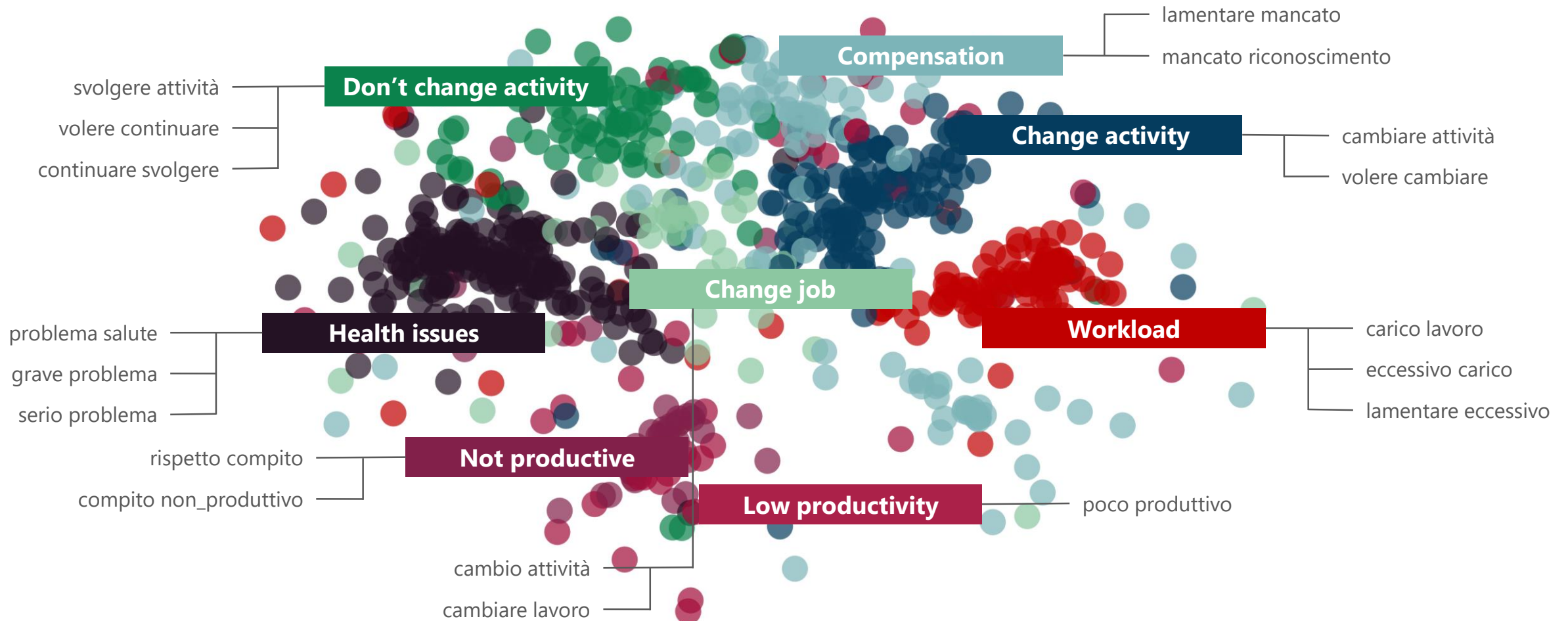
* Changing letters to **lower case**

* **Lemmatization**

* **Typo correction** (110k Italian words + 650 Business-English words¹, Optimal String Align.)

Text clustering results

8 meaningful clusters emerged (k-means + tSNE)





Thank you for your attention!