Why we need a standard data science process

Lesson 1

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Sobering statistics

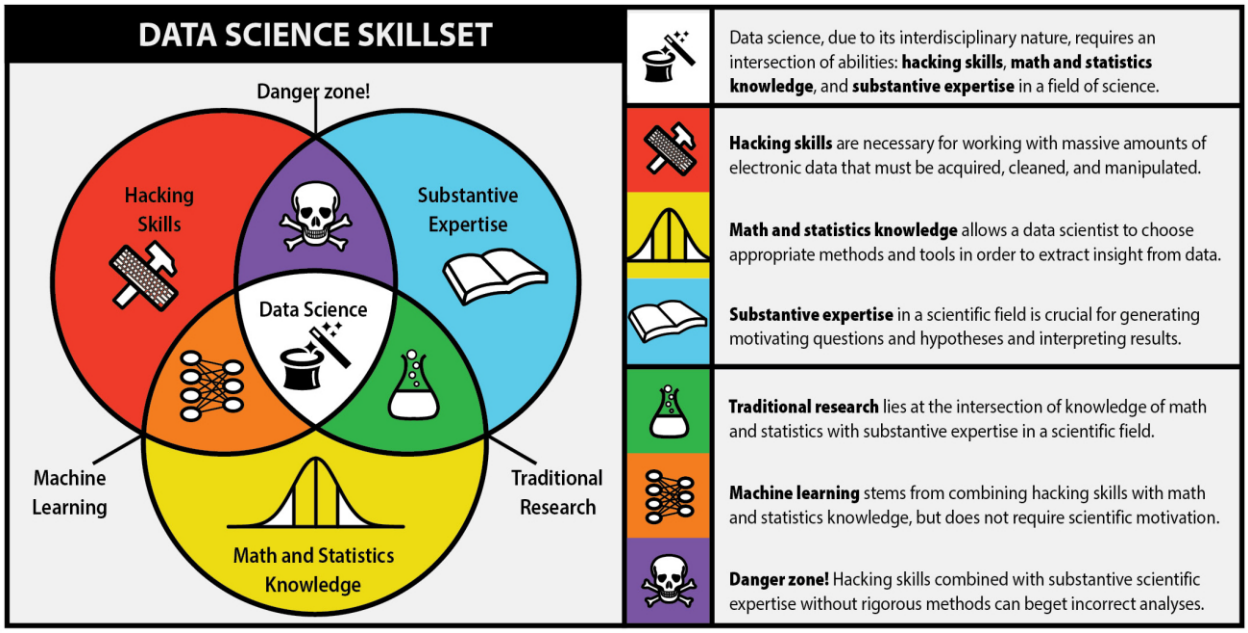
"Only ***27%*** of the big data projects are regarded as successful"

Only ***13%*** of organizations have achieved full-scale production for their Big Data implementations

"Only ***8%*** of the big data projects are regarded as VERY successful"

"Only ***17%*** of survey respondents said they had a well-developed Predictive/Prescriptive Analytics program in place, while 80% said they planned on implementing such a program within five years" – Dataversity 2015 Survey

Source: CapGemini 2014



Source: http://berkeleysciencereview.com/how-to-become-a-data-scientist-before-you-graduate/

How we work: The unspoken challenges of doing data science

Description

Beyond being dubbed "sexiest job in the 21st century," data science is a rewarding career. It's also really hard—not just the technical work itself but also "how to do the work well" in an organization. The term "data scientist" covers a broad range of specific roles ranging from data engineer to data analyst to machine learning expert. Yael Garten explores what data scientists do, how they fit into the broader company organization, and how they can excel at their trade and shares the hard and soft skills required, tips and tricks for success, and challenges to watch out for.

Walking through specific examples, Yael outlines the tips and tactics that she has employed to enable herself and her data science team to thrive, feel empowered, and have impact. She'll leave time for some role playing of challenging scenarios to provide guidance on how to effectively approach and solve some of the difficult issues you might encounter.

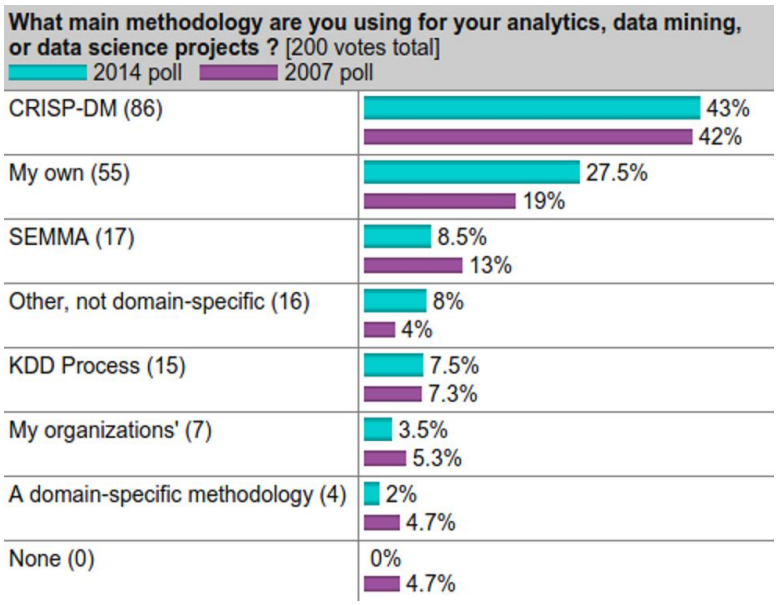
Yael Garten
LinkedIn

Yael Garten leads a team of data scientists at LinkedIn that focuses on understanding and increasing growth and engagement of LinkedIn's 400 million members across mobile and desktop consumer products. Yael is an expert at converting data into actionable product and business insights that impact strategy. Her team partners with

How to Do the Work Well in an Organization?

- We need A Standardized data science process
  - –For data scientists, a standardized data science process can:
    - Support collaboration
    - Support quality assurance in the entire lifecycle of a project
    - Support security control of project assets
    - Support better planning and tracking
  - –For organization, a standardized data science process can:
    - Provide convenience for management to plan and track
    - Accumulate knowledge and expertise over time to continuously improve productivity

Who uses a Process for Data Science?



Source: KDD Nuggets, October 2014

Objectives of a Data Science Process

- For each data science project
  - Standardize the execution of data science project
  - Ensure version control, quality assurance, project asset security control, and collaboration among different personas in the project
  - Boost the productivity
- For the entire data science team
  - Team Learning/Growth. Accumulates knowledge and expertise over time as the evolvement of the standardized git repository structure and the standardized document templates, and the enriched data science utility repository