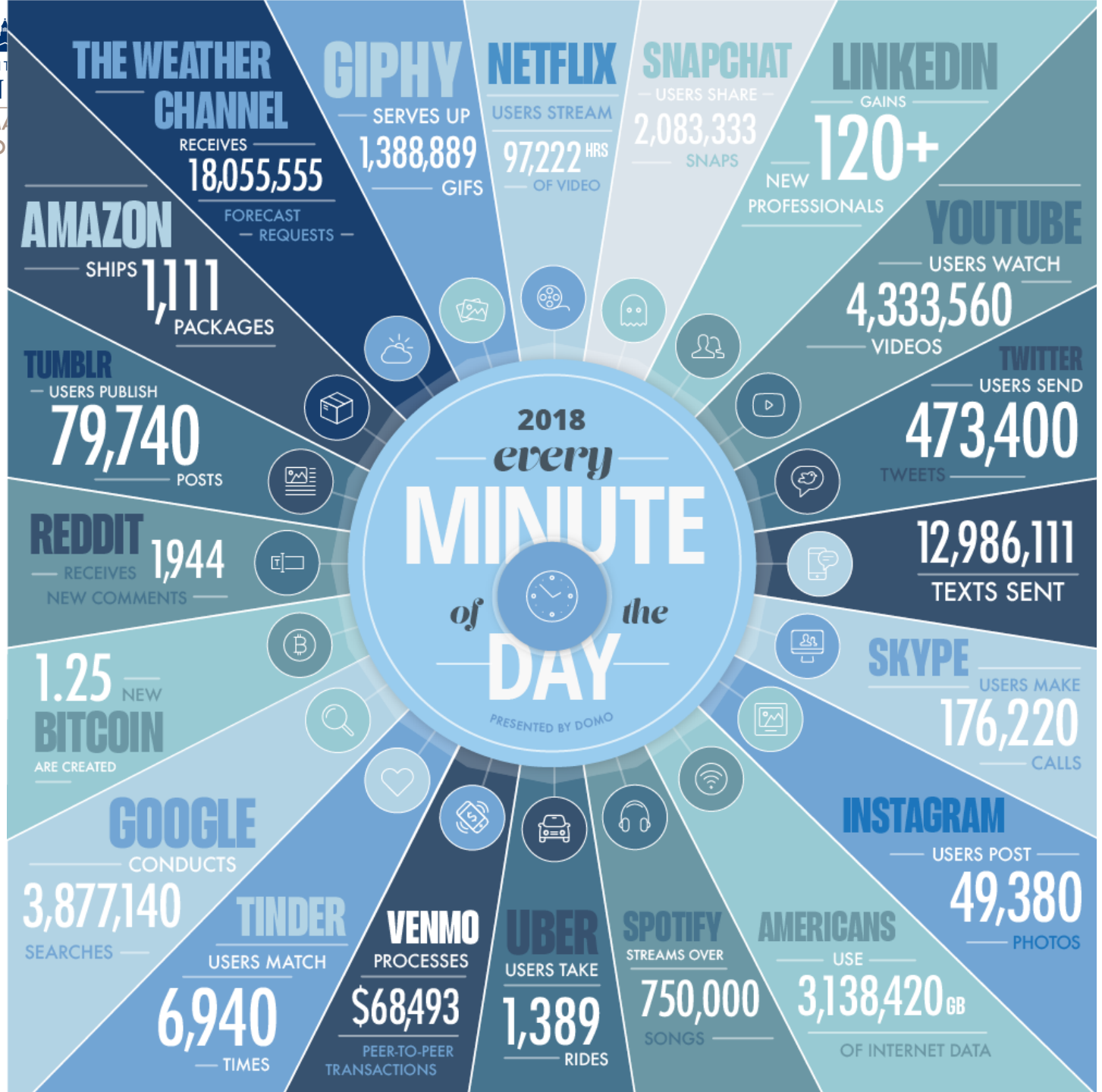
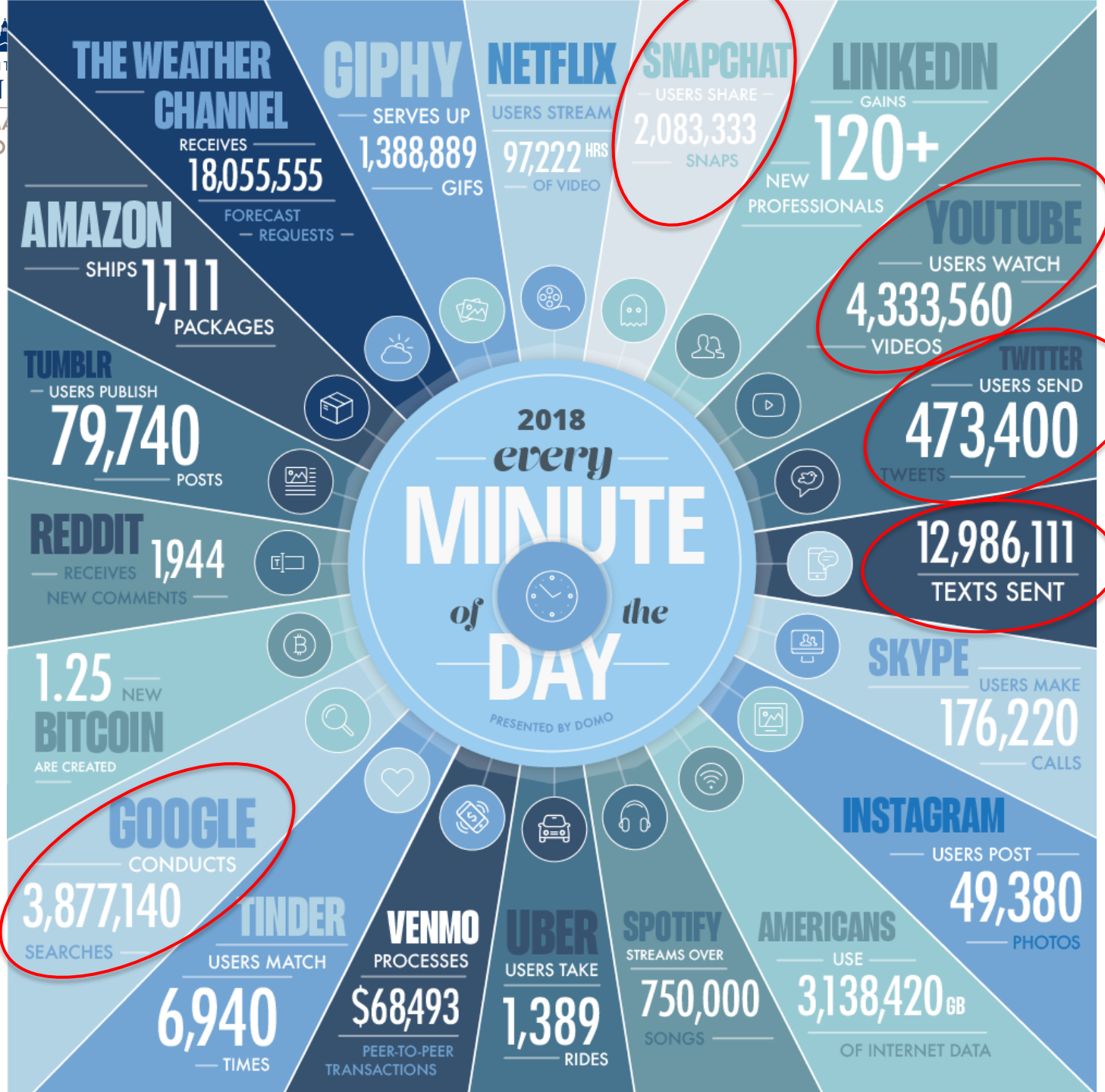
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Introduction à la Data Science

Manuele Kirsch Pinheiro
Luis Angelo Steffenel
Bénédicte Le Grand

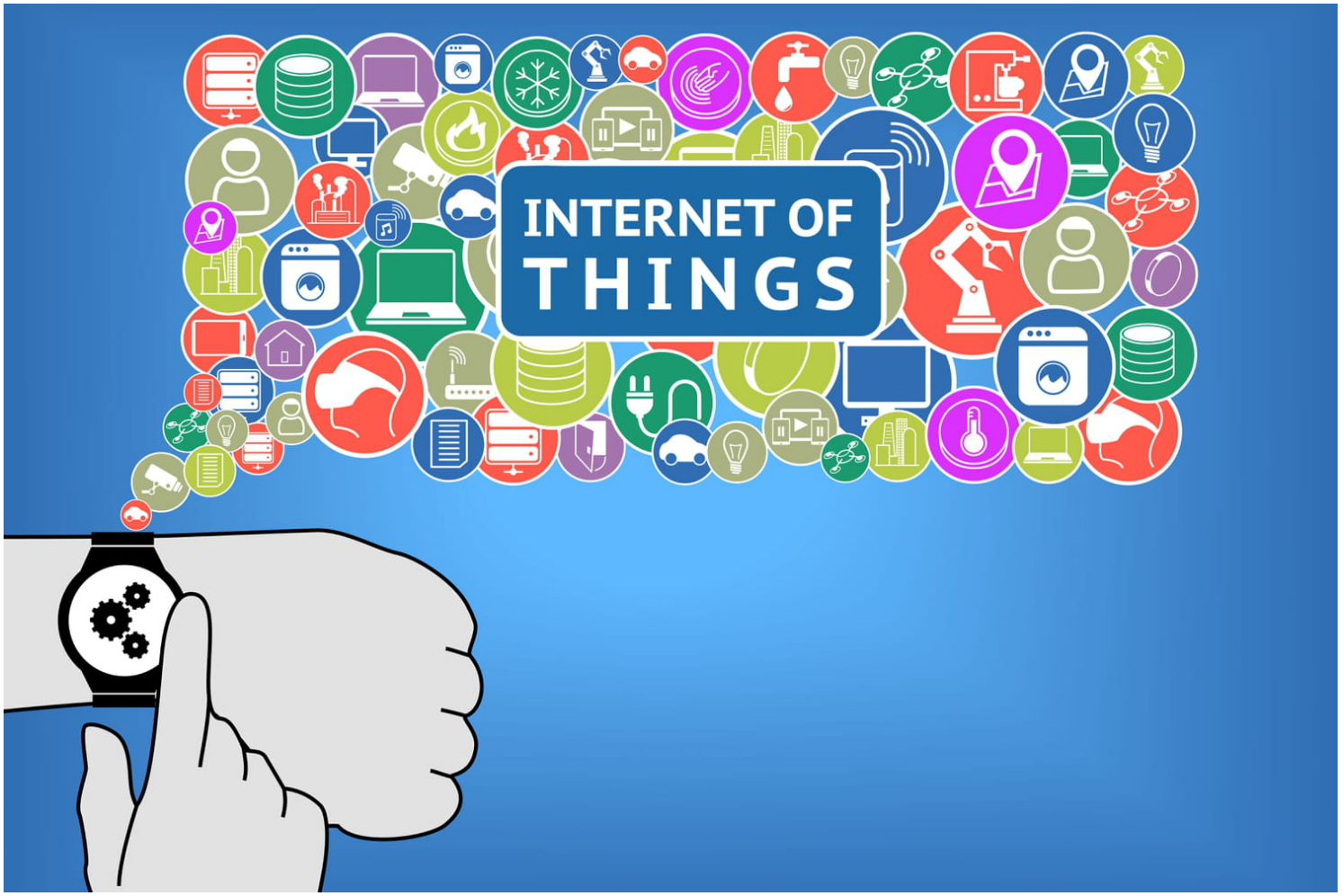




UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Pourquoi ?

Des milliards d'objets connectés



https://www.journaldunet.fr/web-tech/dictionnaire-de-l-iot/1440688-objets-connectes-exemples-d-appareil-et-usages-en-france/

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Internet of Things

Objets capables d'envoyer des données automatiquement :

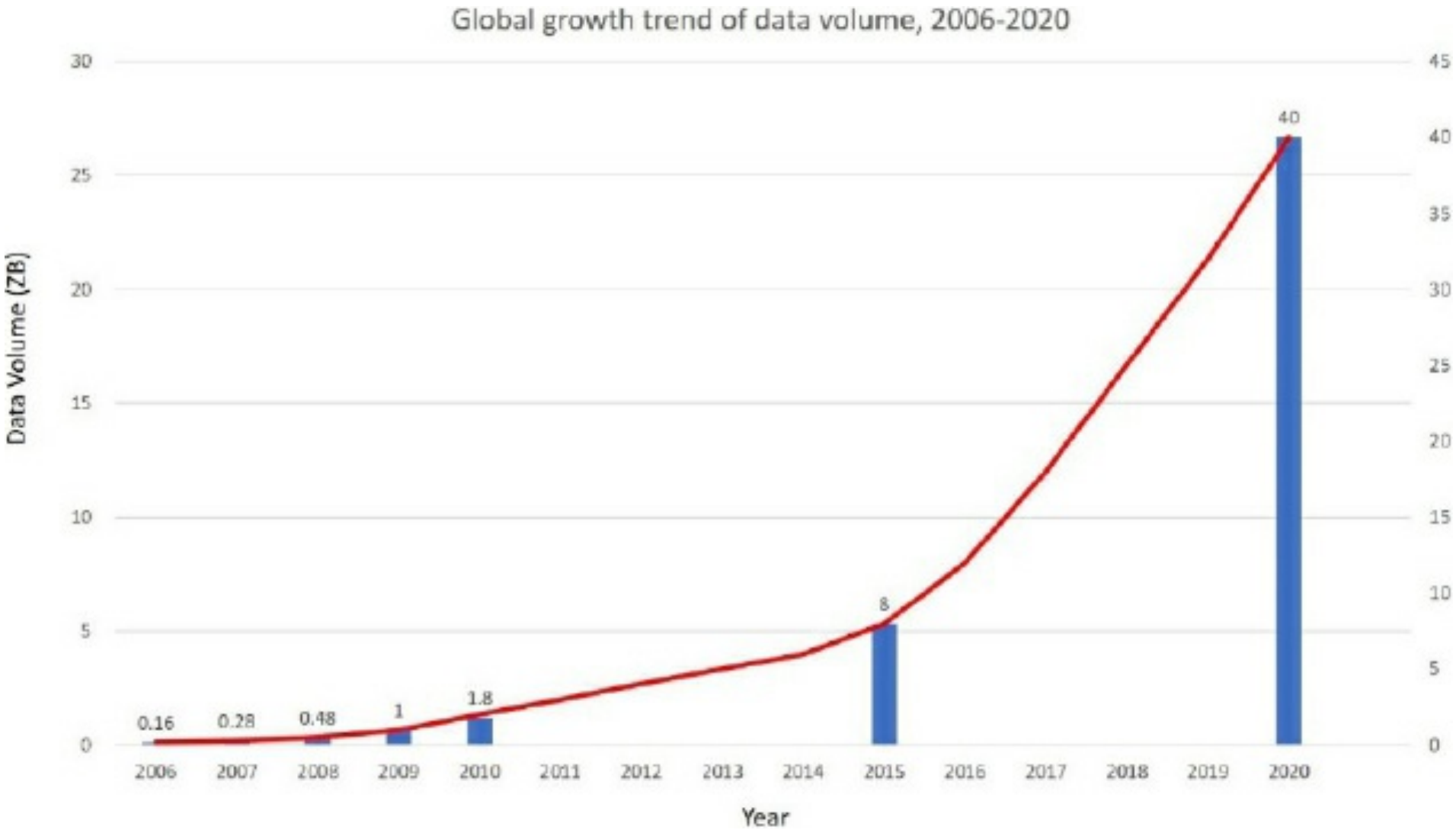
- Voiture avec équipement pour télé-péage,
- Moniteurs de places disponibles dans un parking,
- Moniteur cardiaque implanté chez un humain,
- contrôleurs de la qualité de l'eau,
- Compteur intelligent qui rapporte la consummation d'énergie,
- Détecteur de radiations,
- Traceurs d'objets dans un entrepôt,
- Applis mobiles pour tracer les mouvements et la localisation
- Thermostats intelligents qui ajustent la température des pièces en fonction des prévisions météo et de l'activité dans la maison,
- Équipements domotiques intelligents.

Selon le site statista.com :

- déjà plus de 23 milliards d'équipements IoT aujourd'hui
- prévision : 75 milliards en 2025.

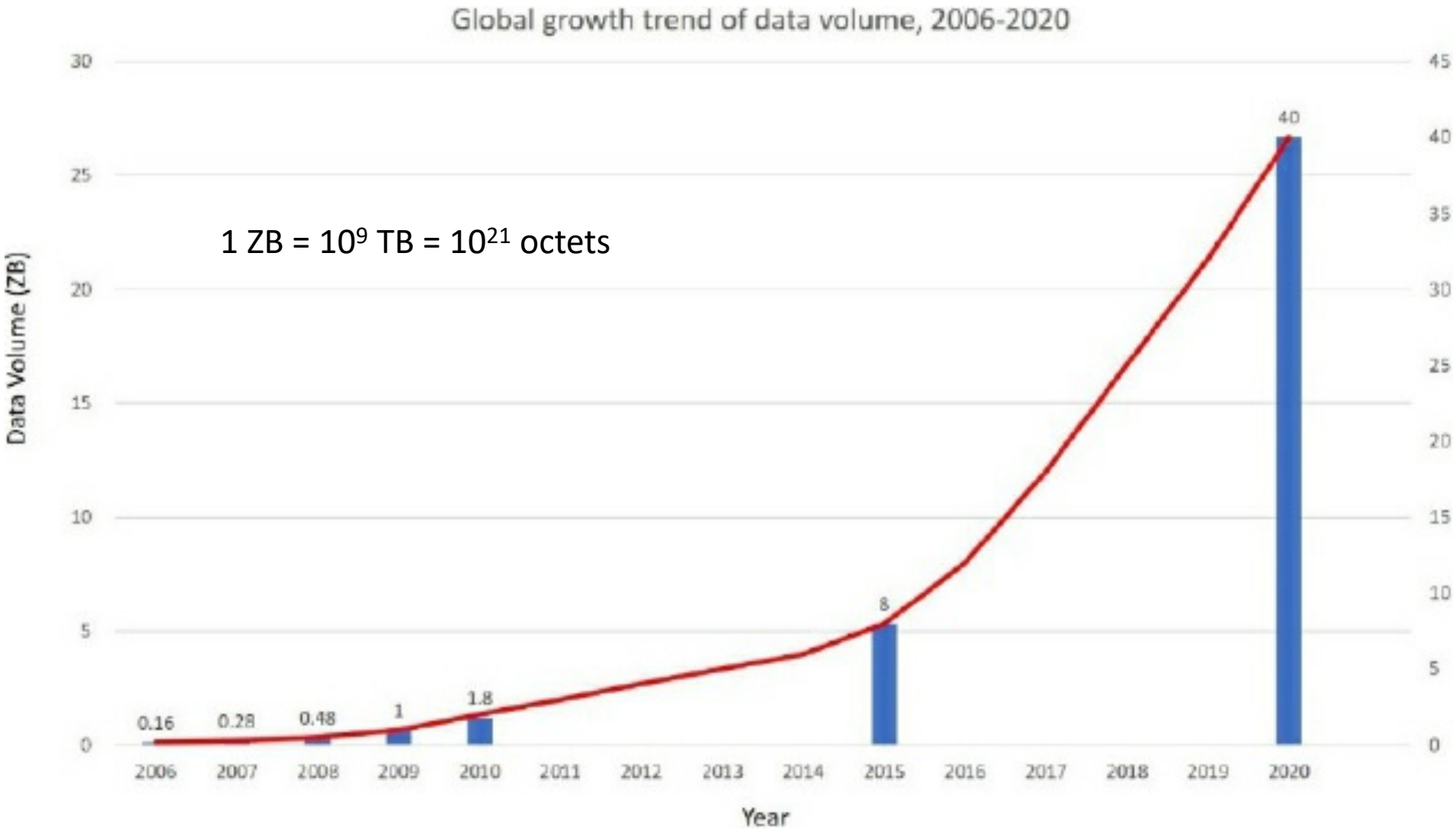
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Croissance du volume de données



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Croissance ~~du volume de données~~ de la valeur des données ???



1 ZB = 10^9 TB = 10^{21} octets

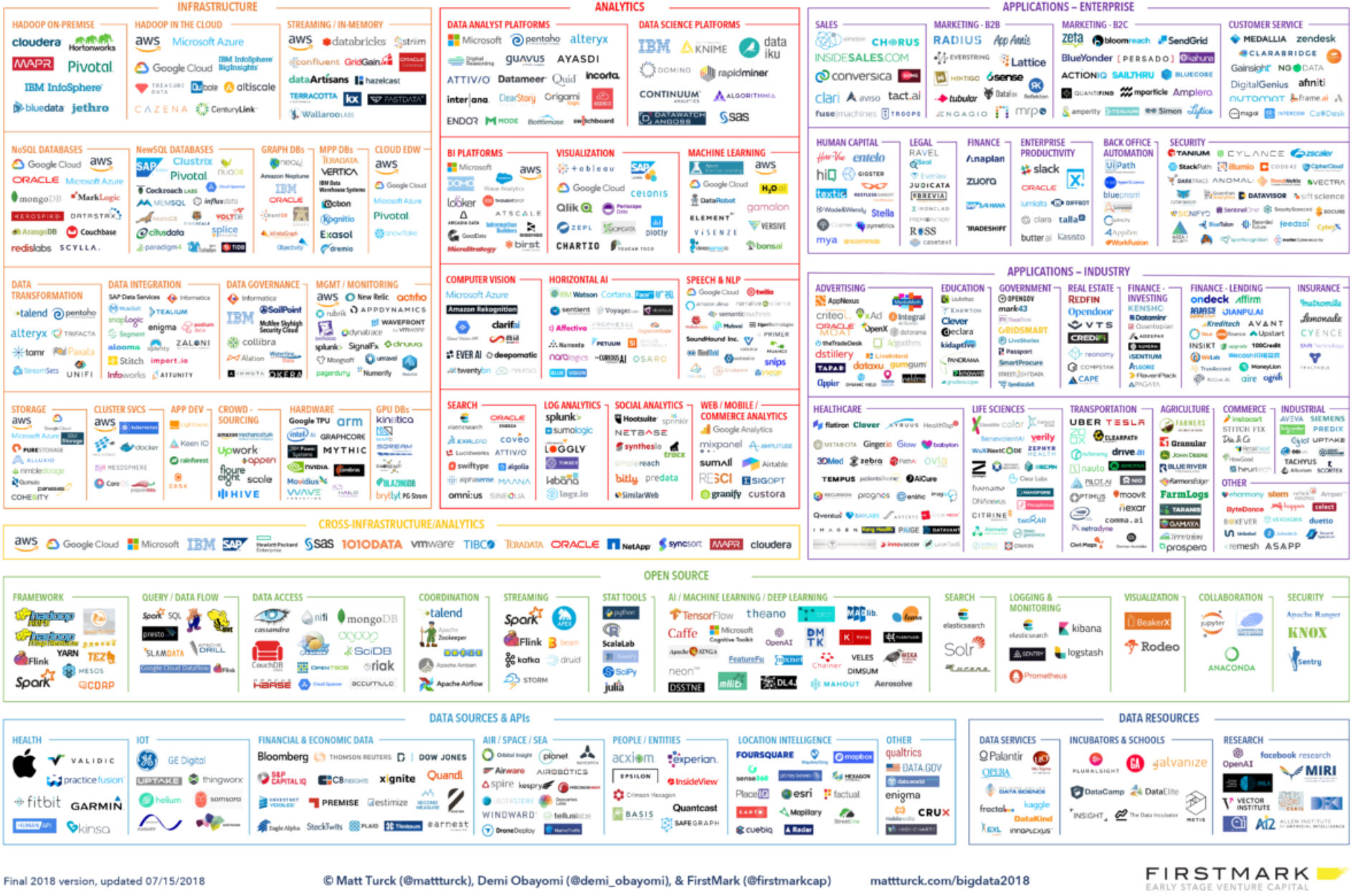
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Big Data

- **Volume** des données générées
- **Vitesse** de production des données
- **Variété** des données
  - Structurées
  - Non structurées
- **V**ariabilité, **V**éracité, **V**alidité, **V**ulnérabilité, **V**olatilité, **V**isualisation, **V**aleur

- Besoin d'avoir des « insights » sur les données
  - On ne sait pas forcément précisément ce que l'on cherche
  - On ne connait pas forcément les éléments importants dans les données

BIG DATA & AI LANDSCAPE 2018



FIRSTMARK
EARLY STAGE VENTURE CAPITAL

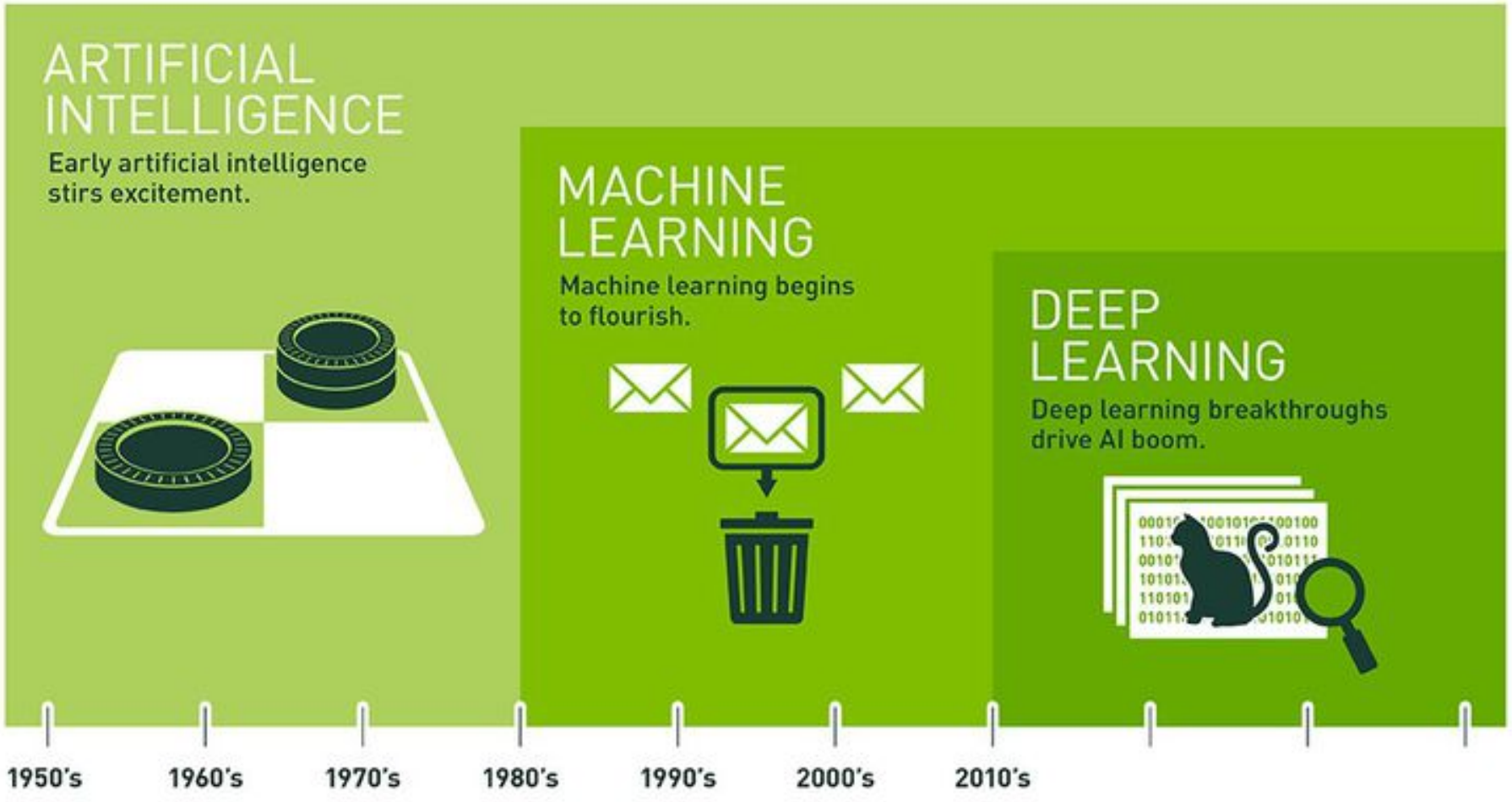
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Data Science

- Applique :
  - des principes scientifiques, des méthodes, des algorithmes et des processus
  - pour extraire des connaissances, de l'information
  - en collectant, traitant et analysant des données structurées et non structurées
    - Sources de données structurées : par exemple SGBD (Oracle, MySQL,…)
    - Sources de données non structurées : texte, audio, vidéo, documents.

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

IA, Machine Learning, Deep Learning ?

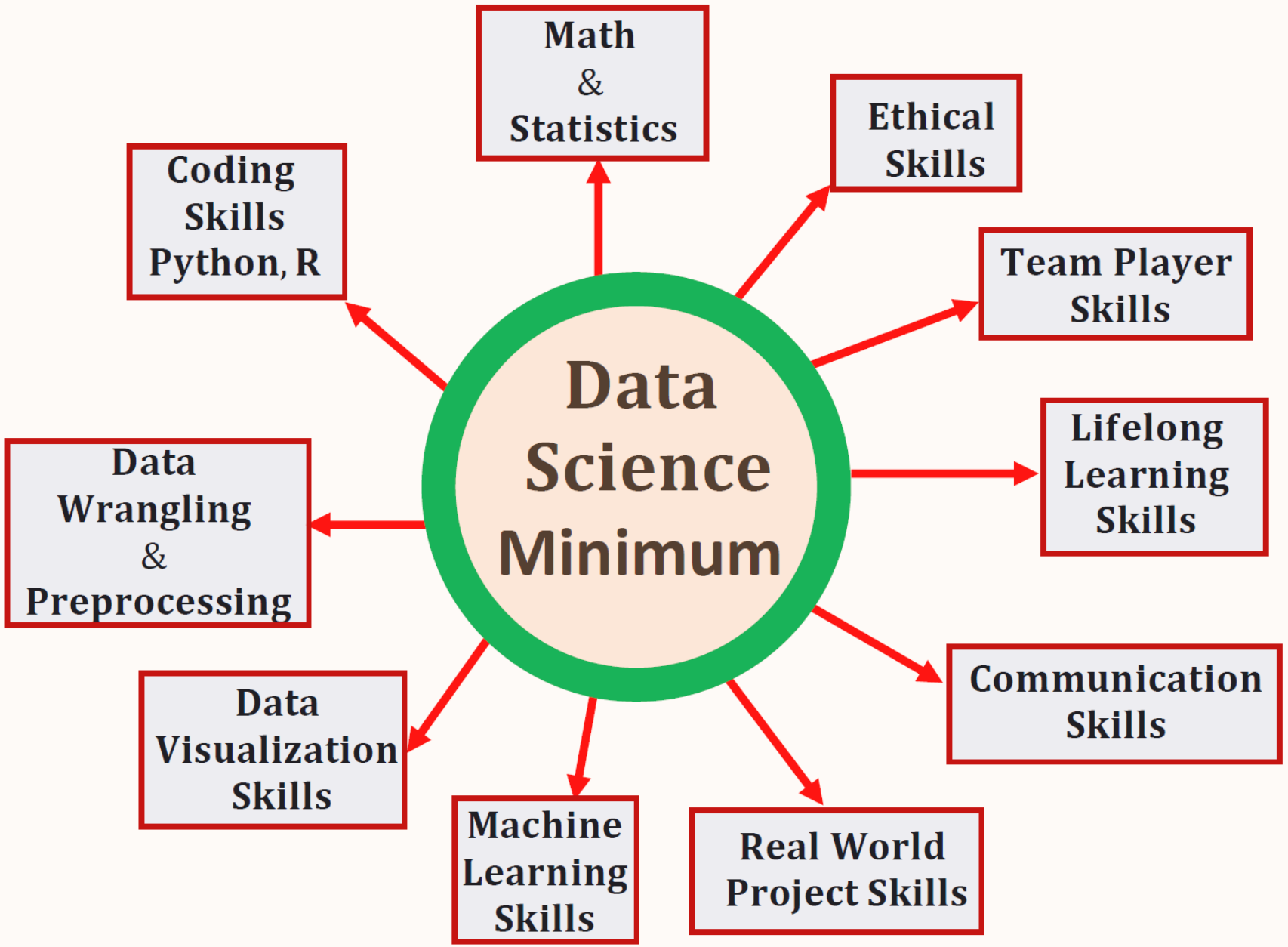


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

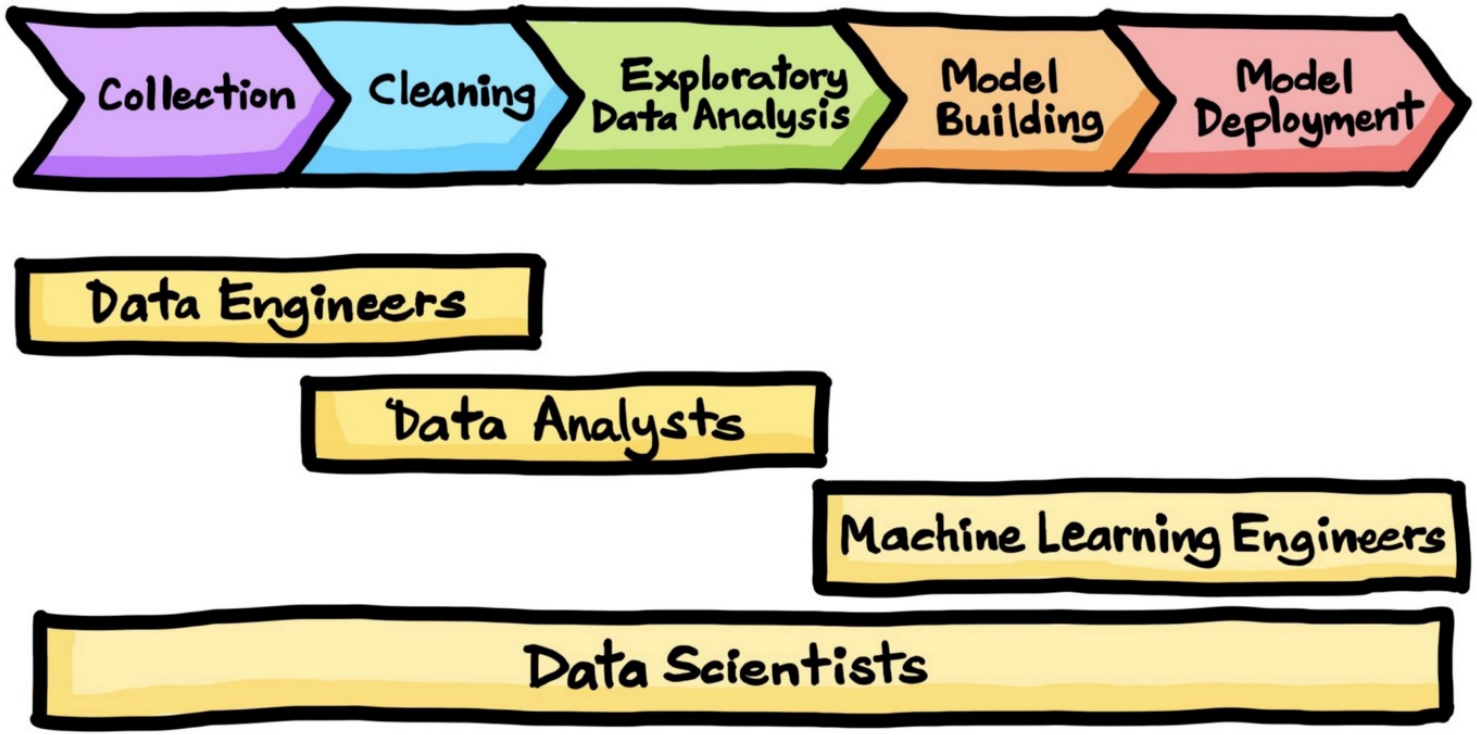
Domaines d'application

- Marketing, relation client, systèmes de recommandation
- Santé / médecine
- Secteur des banques et assurances, détection de fraudes
- Ressources humaines
- Cybersécurité
- Reconnaissance vocale/faciale, assistants personnels
- Prévisions de trafic, météo
- Analyse de réseaux sociaux, détection de tendances
- Maintenance prédictive
- …

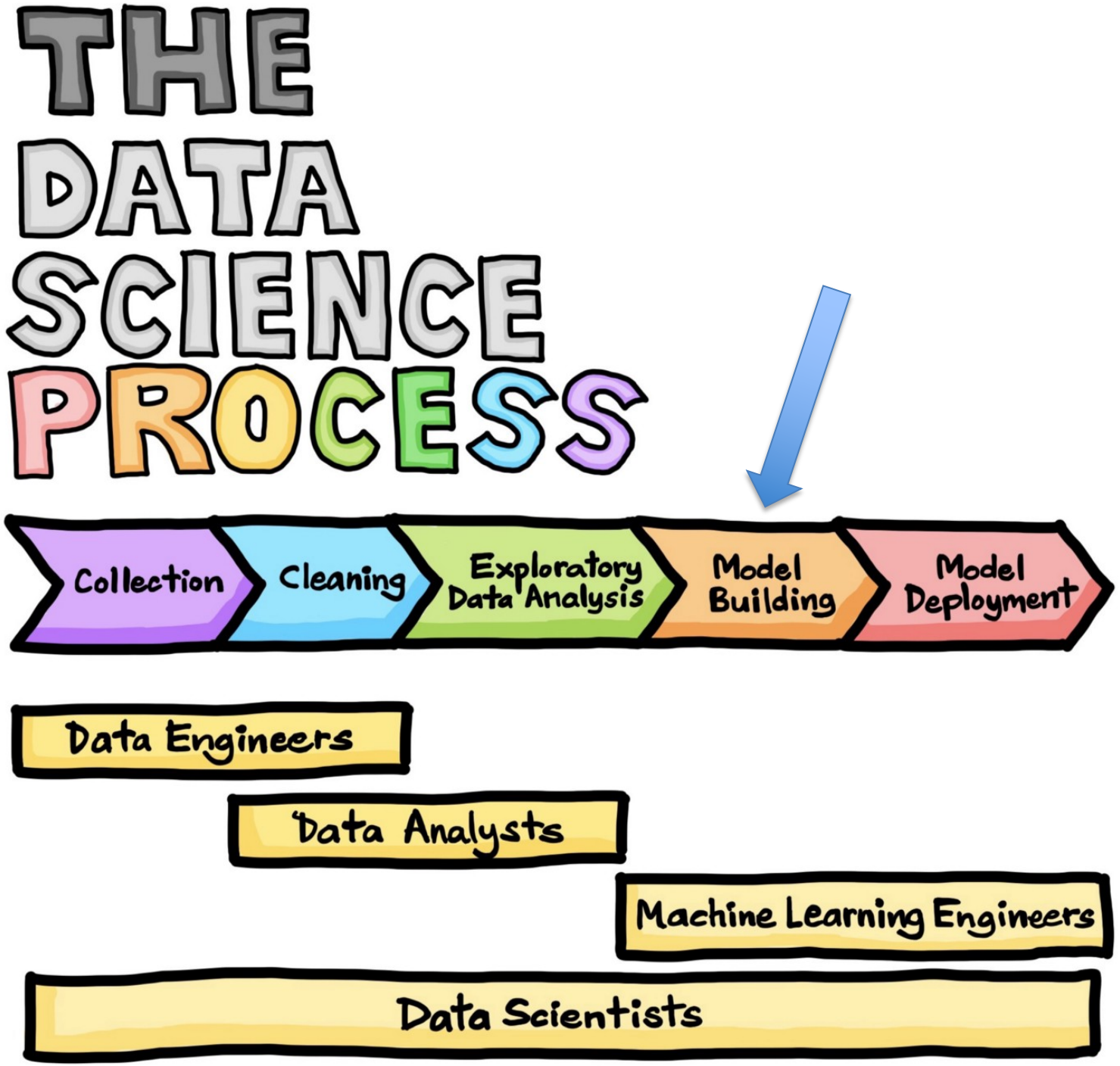


https://towardsdatascience.com/data-science-minimum-10-essential-skills-you-need-to-know-to-start-doing-data-science-e5a5a9be5991

THE DATA SCIENCE PROCESS



https://towardsdatascience.com/the-data-science-process-a19eb7ebc41b

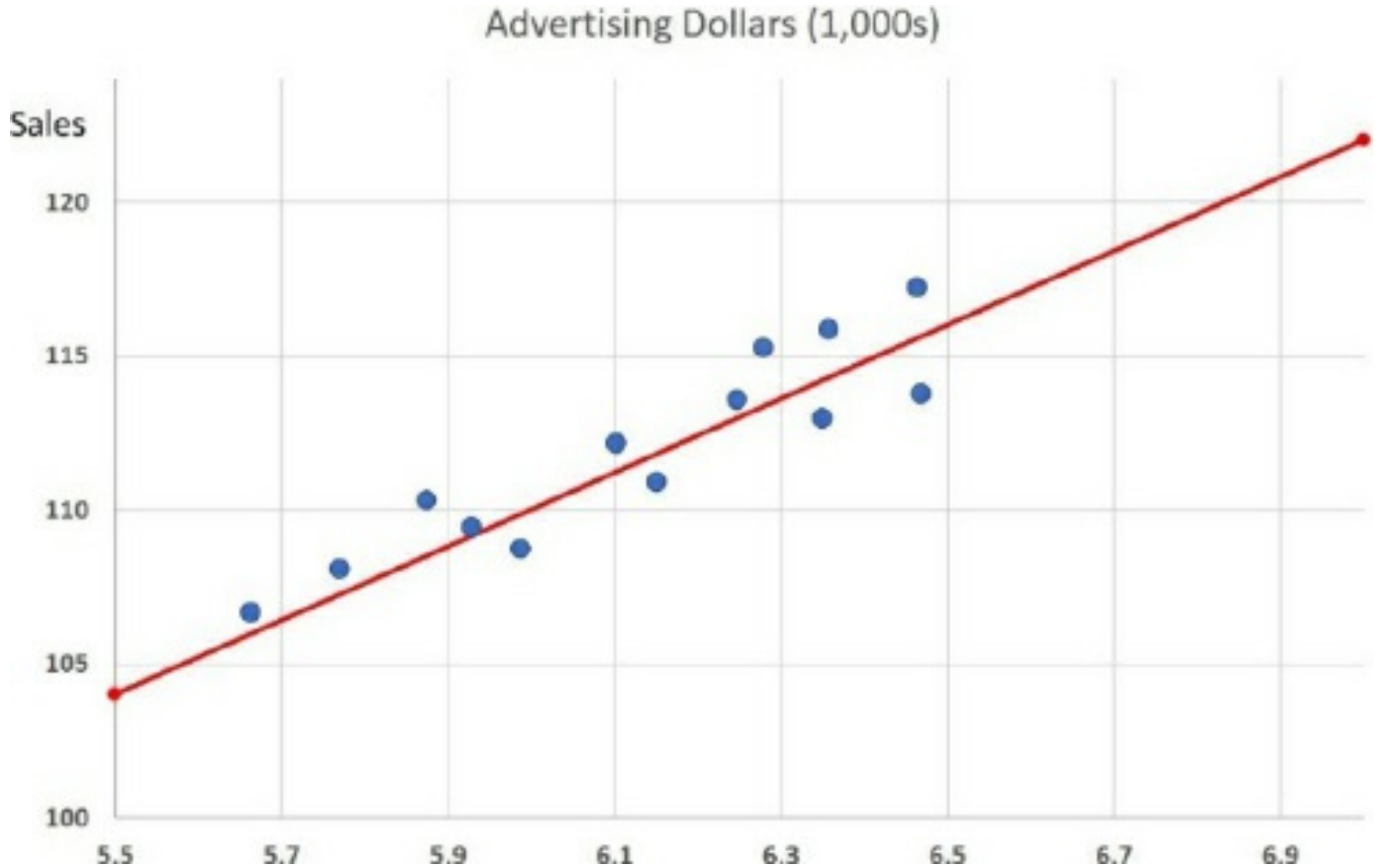


https://towardsdatascience.com/the-data-science-process-a19eb7ebc41b

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Construction d'un modèle

- Modèle = ensemble d'hypothèses à propos des données
- Exemple de modèle : relation quasi linéaire entre le volume des ventes et la somme dépensée en publicité



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

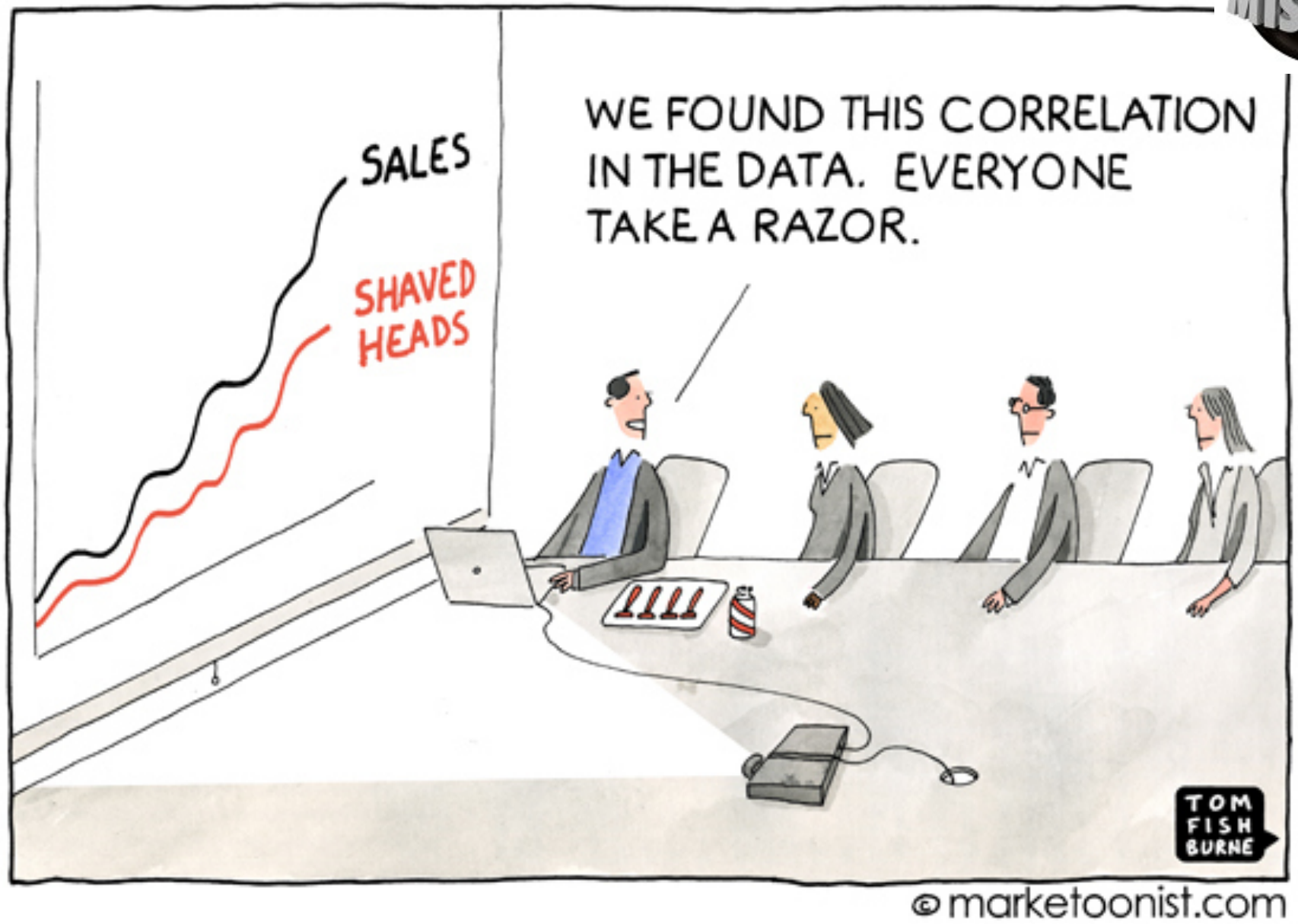
Importance d'avoir un « bon » modèle

- Pour faire des hypothèses / prédictions correctes

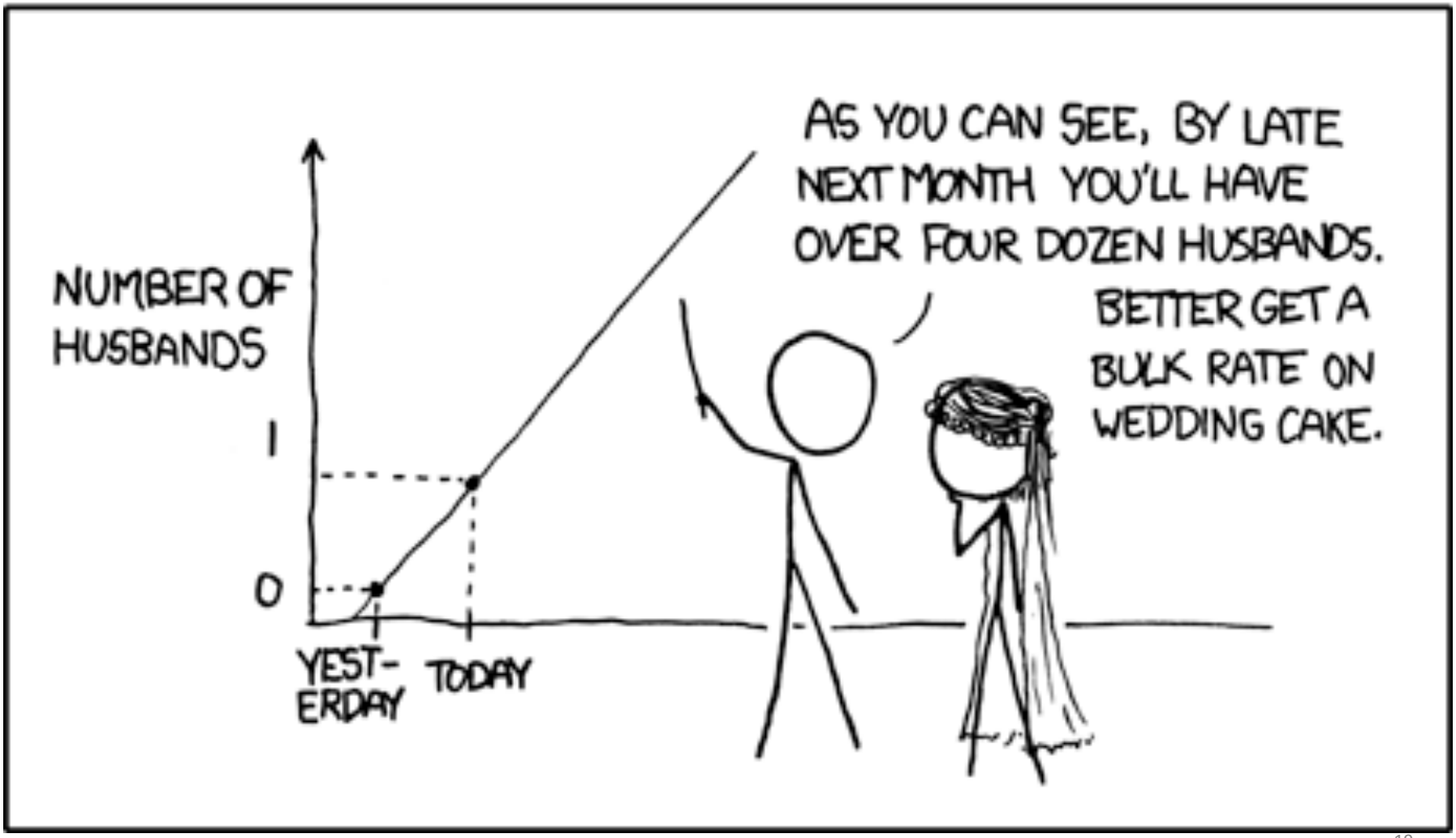
- Quel algorithme choisir ?

- Comment évaluer la qualité du modèle construit ?


BIG
MISTAKE!

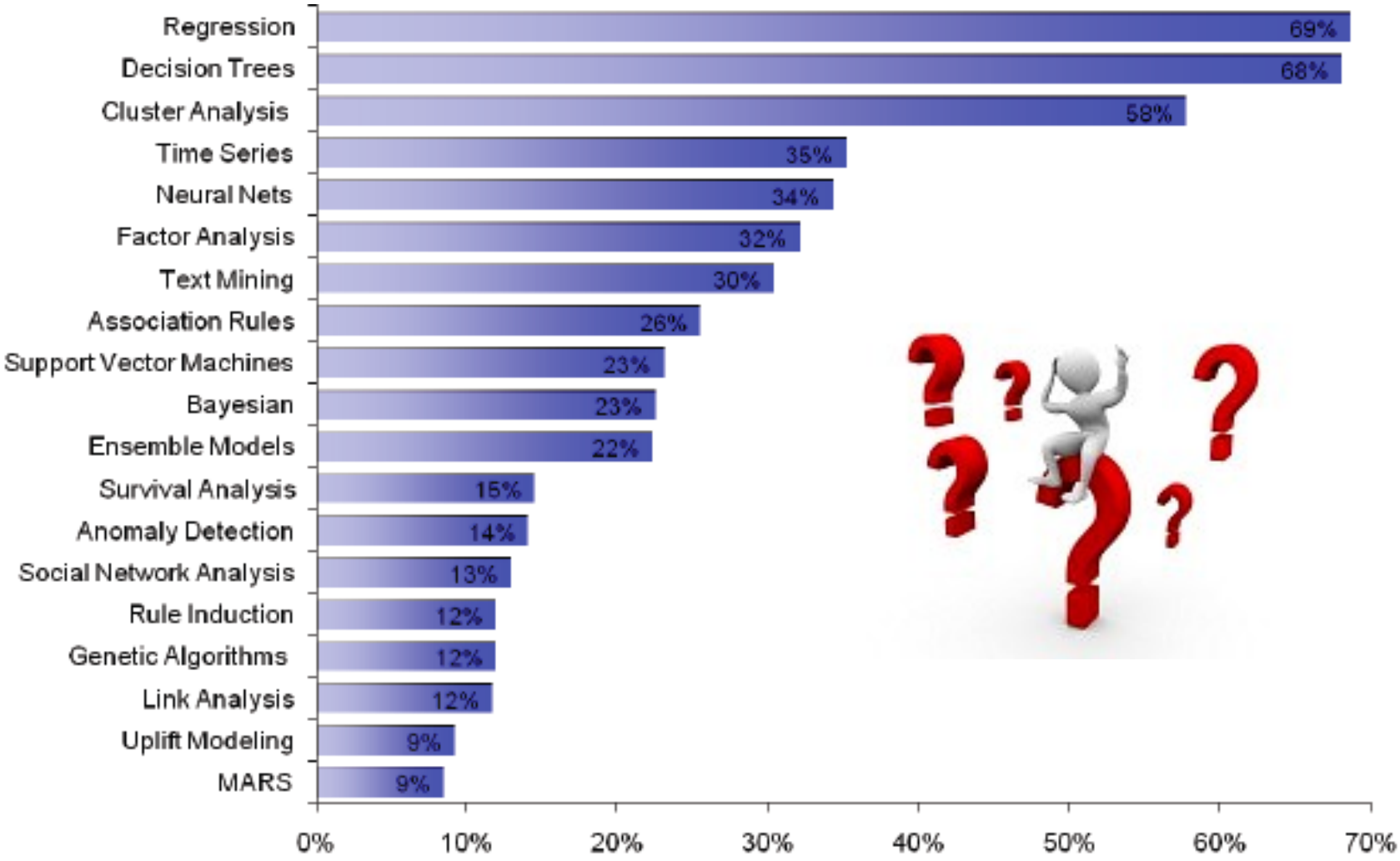

SALES
SHAVED HEADS
WE FOUND THIS CORRELATION IN THE DATA. EVERYONE TAKE A RAZOR.
TOM FISH BURNE
©marketoonist.com

MY HOBBY: EXTRAPOLATING



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

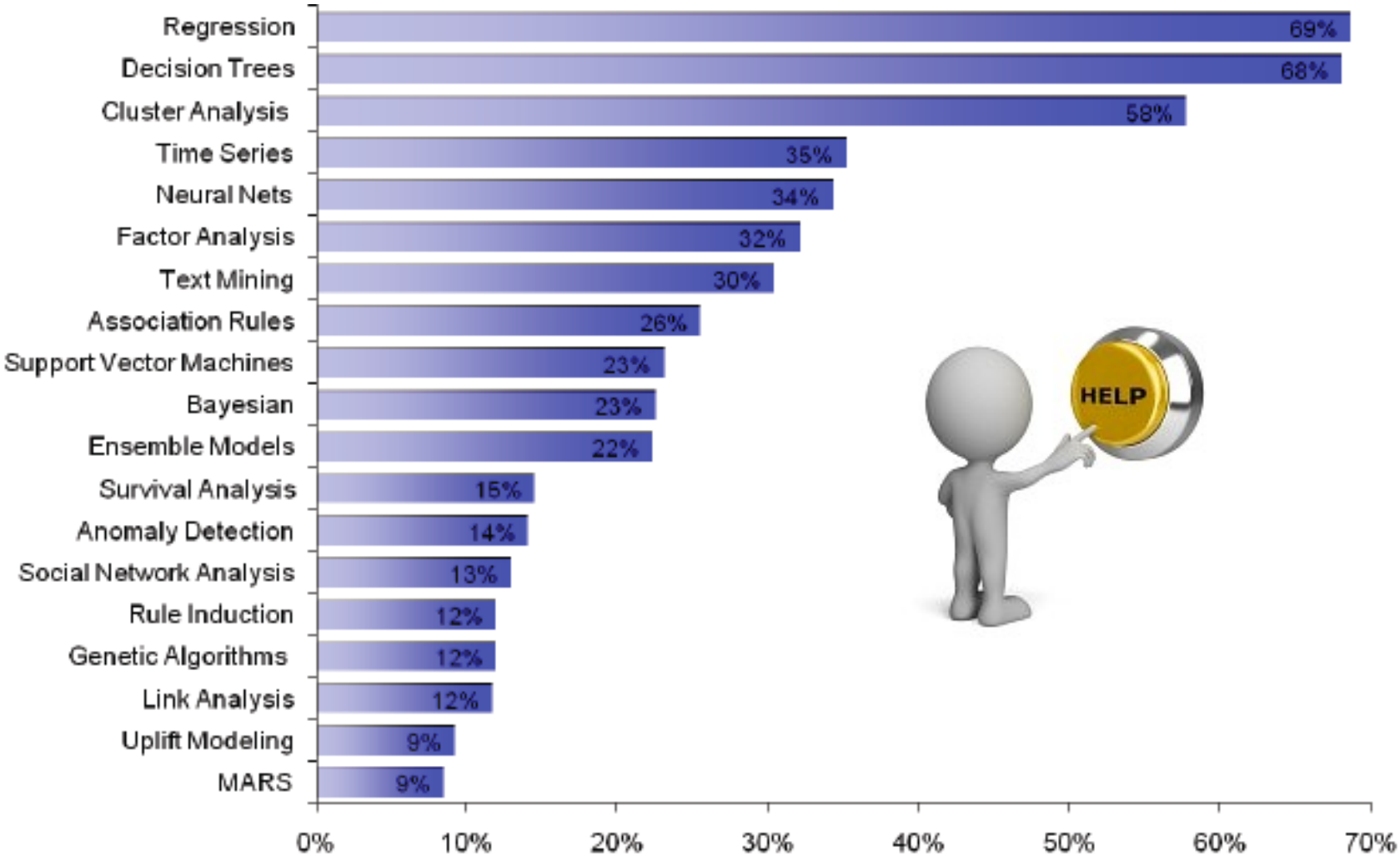
Quel algorithme utiliser ?



Question: What algorithms/analytic methods do you TYPICALLY use? (Select all that apply)

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Quel algorithme utiliser ?

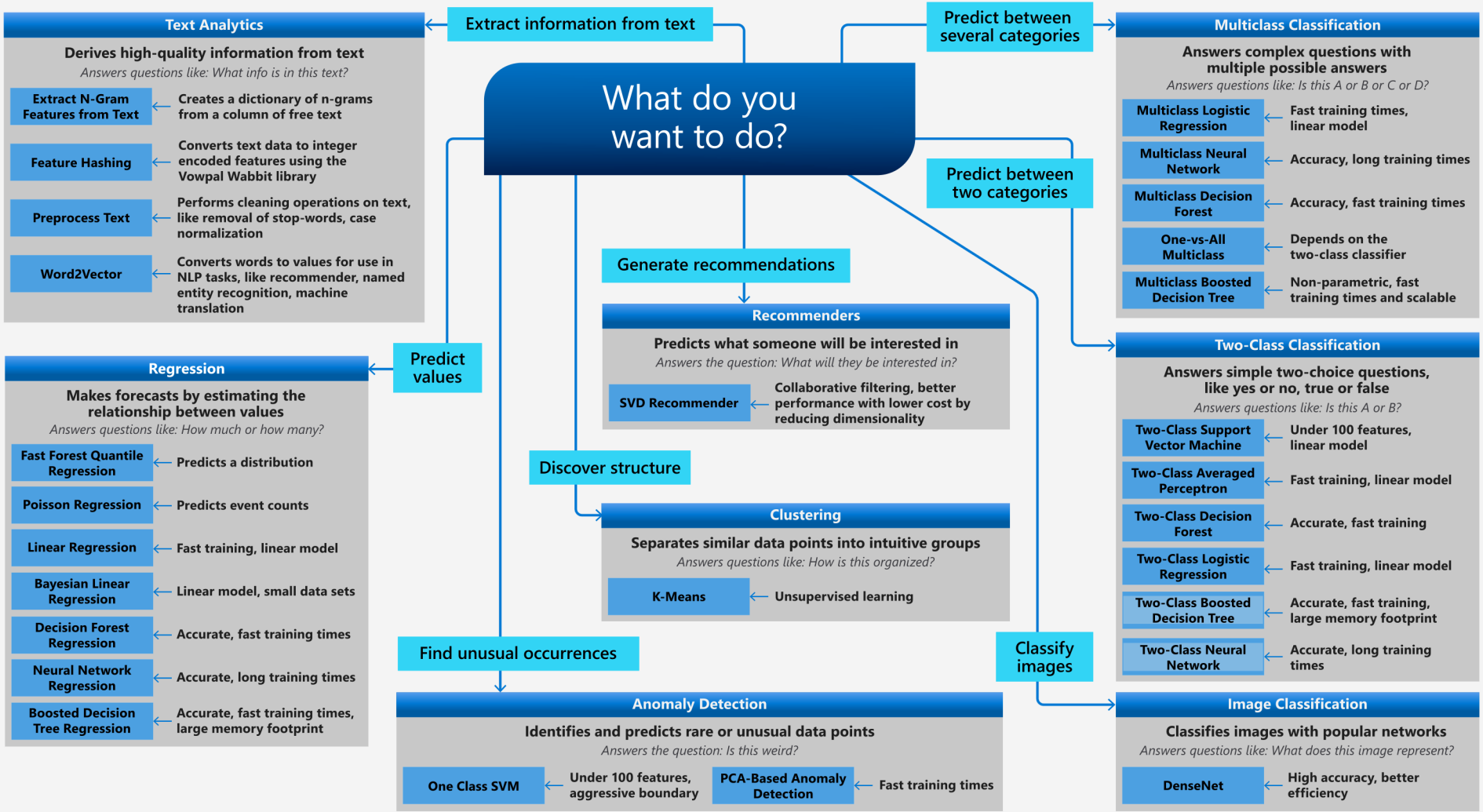


Question: What algorithms/analytic methods do you TYPICALLY use? (Select all that apply)

https://s3.amazonaws.com/MLMastery/MachineLearningAlgorithms.png

Microsoft Azure Machine Learning Algorithm Cheat Sheet

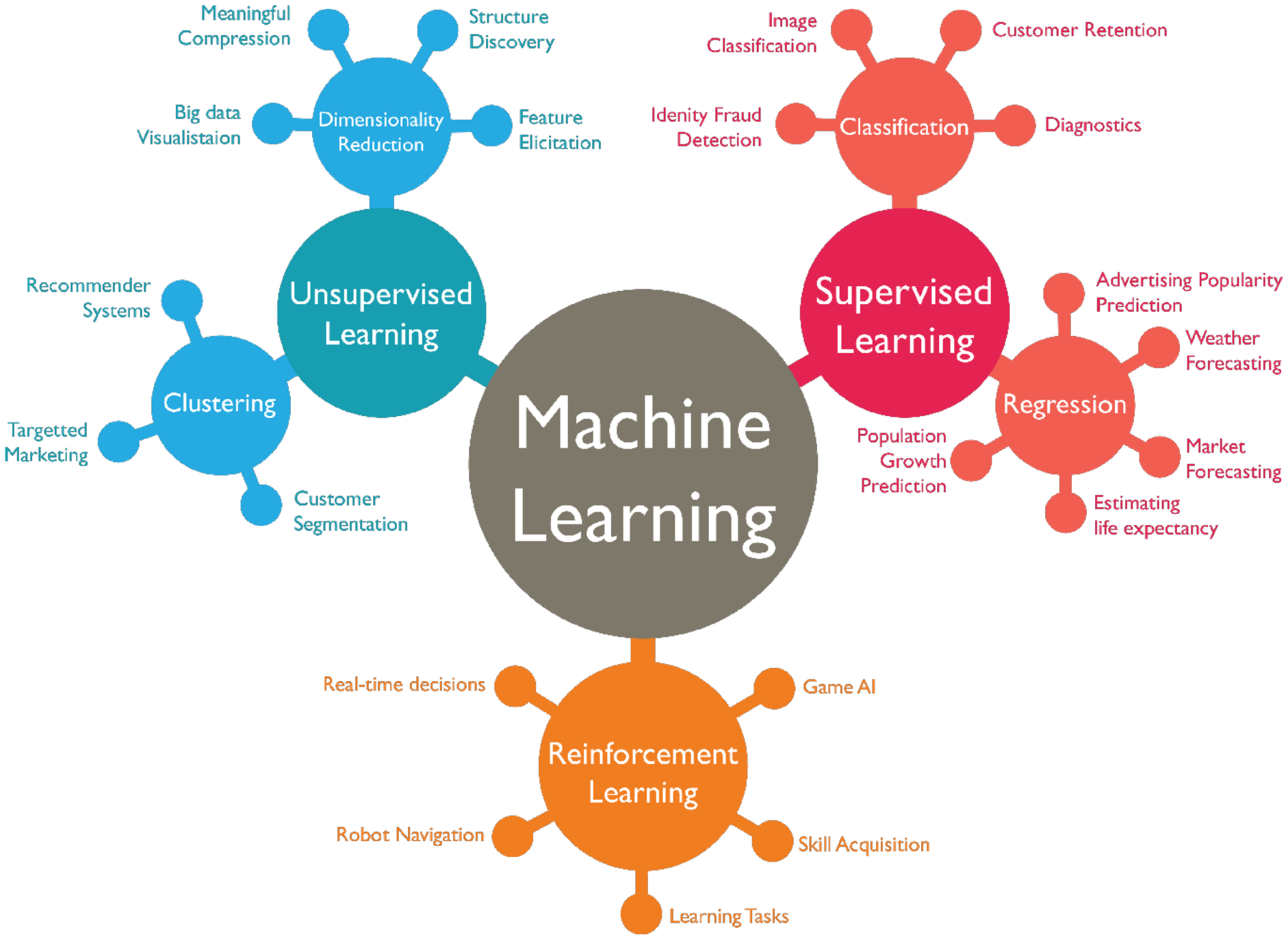
This cheat sheet helps you choose the best machine learning algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the goal you want to achieve with your data.



 Share this poster: aka.ms/mlcheatsheet

Microsoft

https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet



https://wordstream-files-prod.s3.amazonaws.com/s3fs-public/machine-learning.png

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

UNIVERS
PANTHÉON
ÉCOLE DE
DE LA S

THE DATA SCIENCE PROCESS

Bienvenue dans l'aventure !!!

https://towardsdatascience.com/the-data-science-process-a19eb7ebc41b