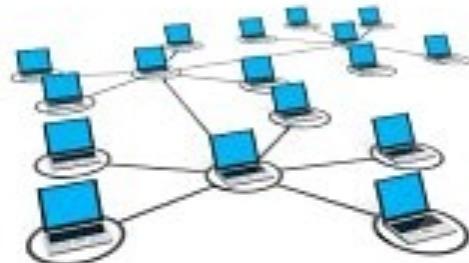


Introduction à la Data Science et au Machine Learning



Manuele Kirsch Pinheiro
Luiz Angelo Steffenel
Bénédicte Le Grand

THE WEATHER CHANNEL

RECEIVES —

18,055,5

AMAZON
SHIPS 1,111
PACKAGES

FORECAST
— REQUESTS —

TUMBLR
— USERS PUBLISH
79,740
— POSTS

REDDIT — RECEIVES **1,944** NEW COMMENTS —

1.25 NEW
BITCOIN
ARE CREATED

GOOGLE
CONDUCTS
3,877,140
SEARCHES

3,877,140
SEARCHES —
6,940
— TIMES

GIPHY
SERVES UP
1,388,889
GIFS

NETFLIX
USERS STREAM
97,222 HRS
— OF VIDEO

LINKEDIN GAINS **120+**

ONALS — YOUTUBE
— USERS WATCH
4,333,560
— VIDEOS — TWITTER

VIDEOS
TWITTER
USERS SEND
473,400
TWEETS

12,986,111
TEXTS SENT

SKYPE USERS MAKE
176,220 CALLS

INSTAGRAM
— USERS POST —
49,380
— PHOTOS
— CANVAS USE —
138,420 GB
— OF INTERNET DATA

2018 every **MINUTE** of **DAY** the

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 consommation 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 :

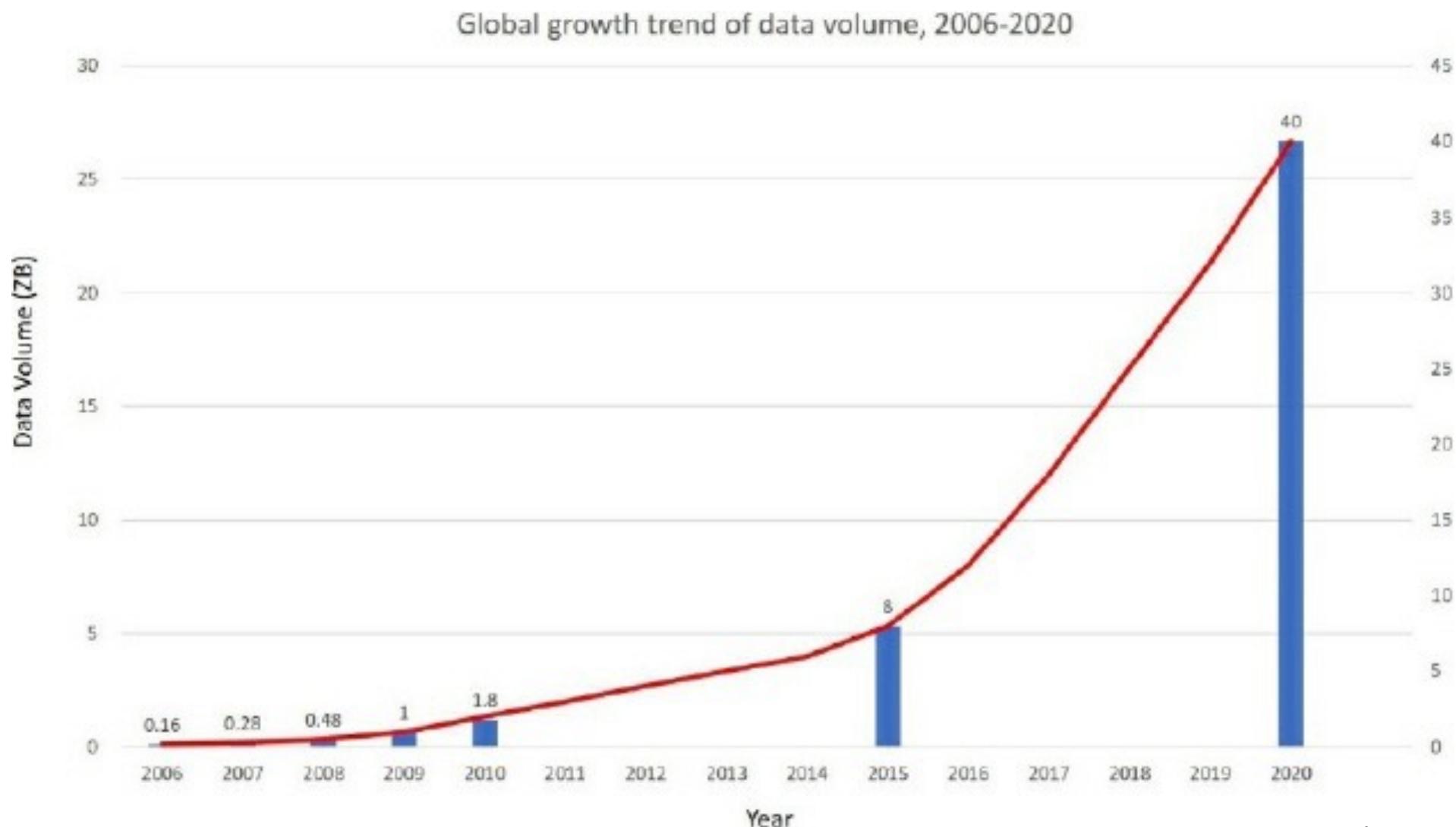
- déjà plus de 23 milliards d'équipements IoT aujourd'hui
- prévision : 75 milliards en 2025.



UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Croissance du volume de données





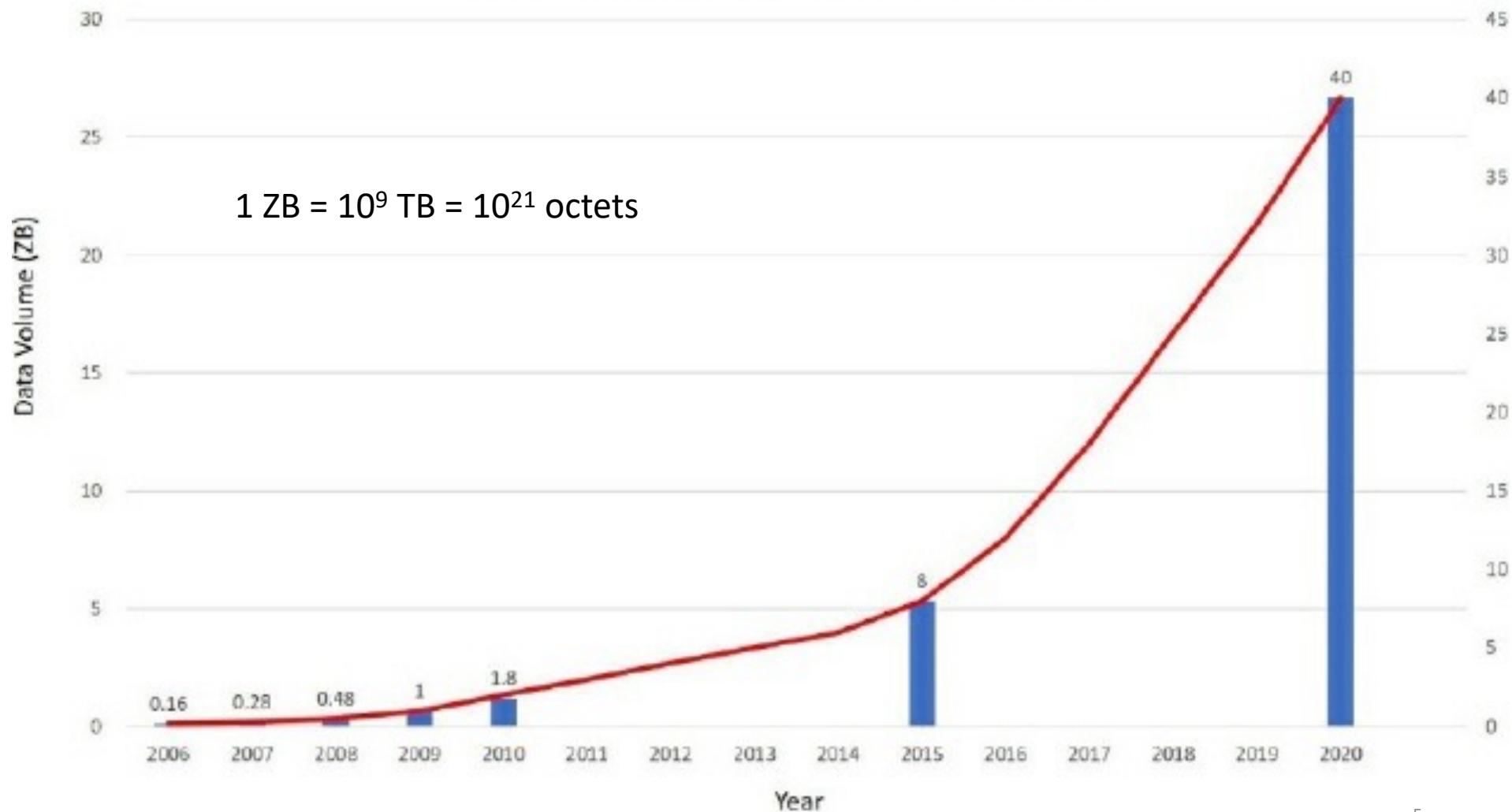
UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

ÉCOLE DE MANAGEMENT
DE LA SORBONNE

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

Global growth trend of data volume, 2006-2020





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
- **Variabilité, Véracité, Validité, Vulnérabilité, Volatilité, Visualisation, Valeur**
- Besoin d'avoir des « insights » sur les données
 - On ne sait pas forcément précisément ce que l'on cherche
 - On ne connaît pas forcément les éléments importants dans les données



BIG DATA & AI LANDSCAPE 2018





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.



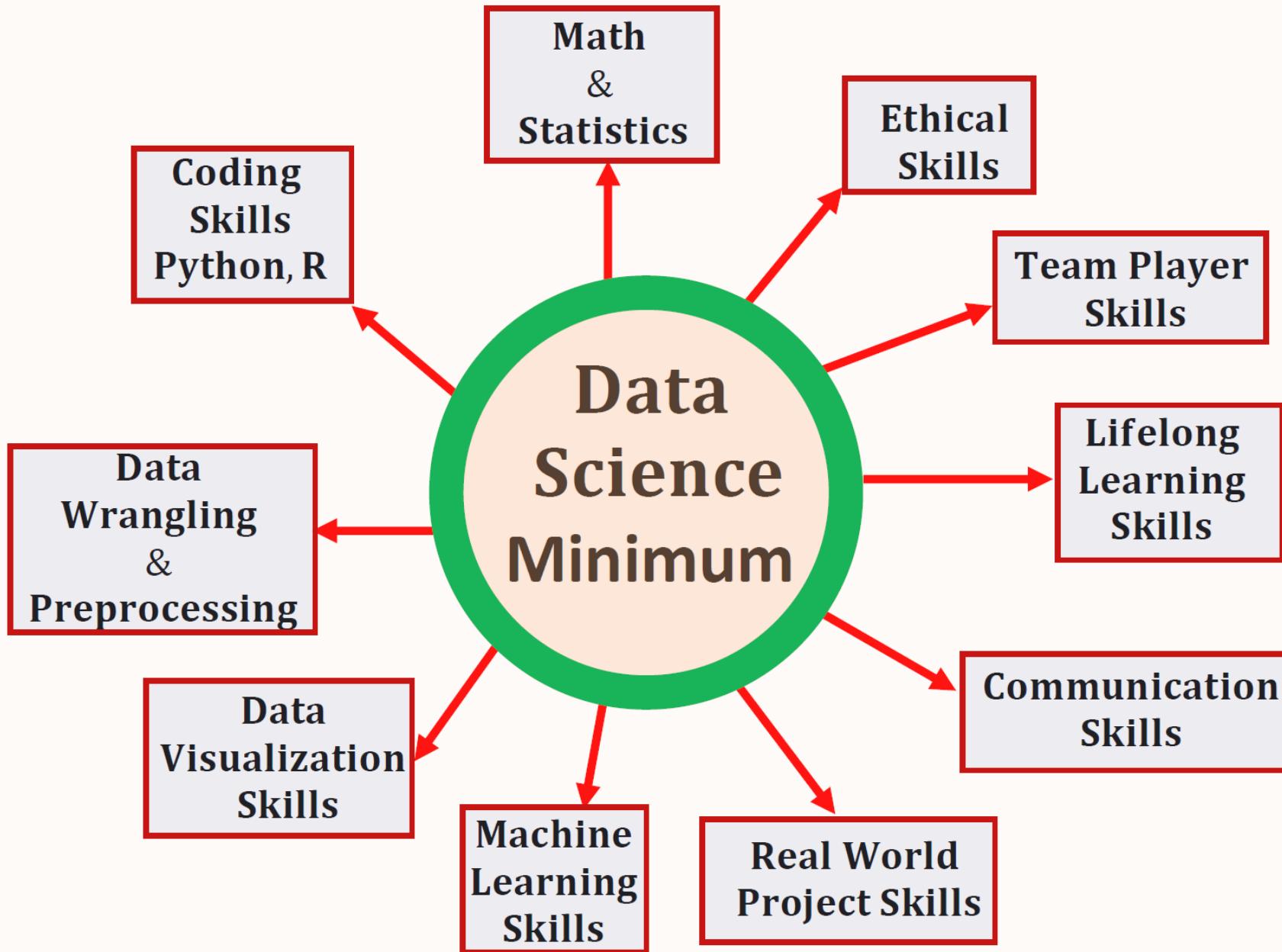
UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

ÉCOLE DE MANAGEMENT
DE LA SORBONNE

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
- ...





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

ÉCOLE DE MANAGEMENT
DE LA SORBONNE

IA, Machine Learning, Deep Learning ?

ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



MACHINE LEARNING

Machine learning begins to flourish.



DEEP LEARNING

Deep learning breakthroughs drive AI boom.



1950's

1960's

1970's

1980's

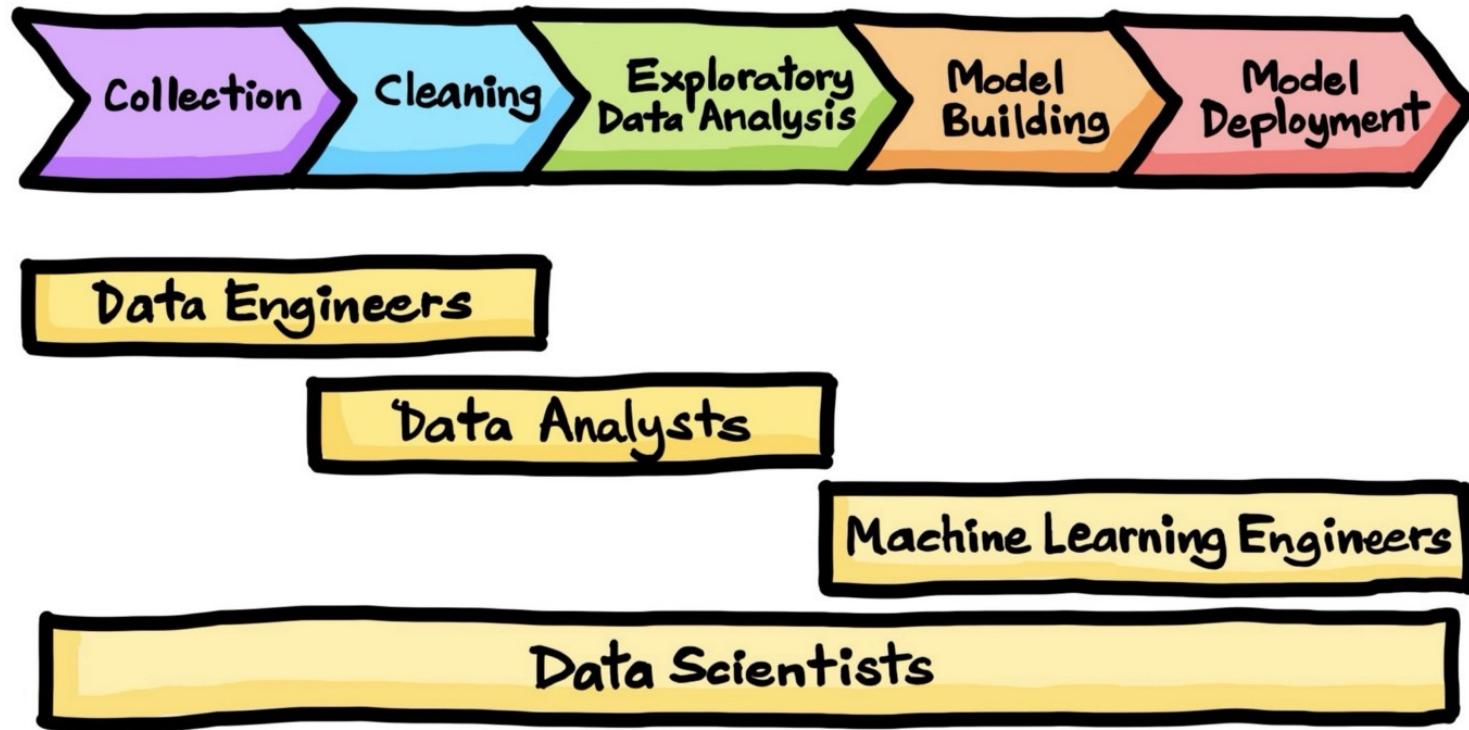
1990's

2000's

2010's

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.

THE DATA SCIENCE PROCESS





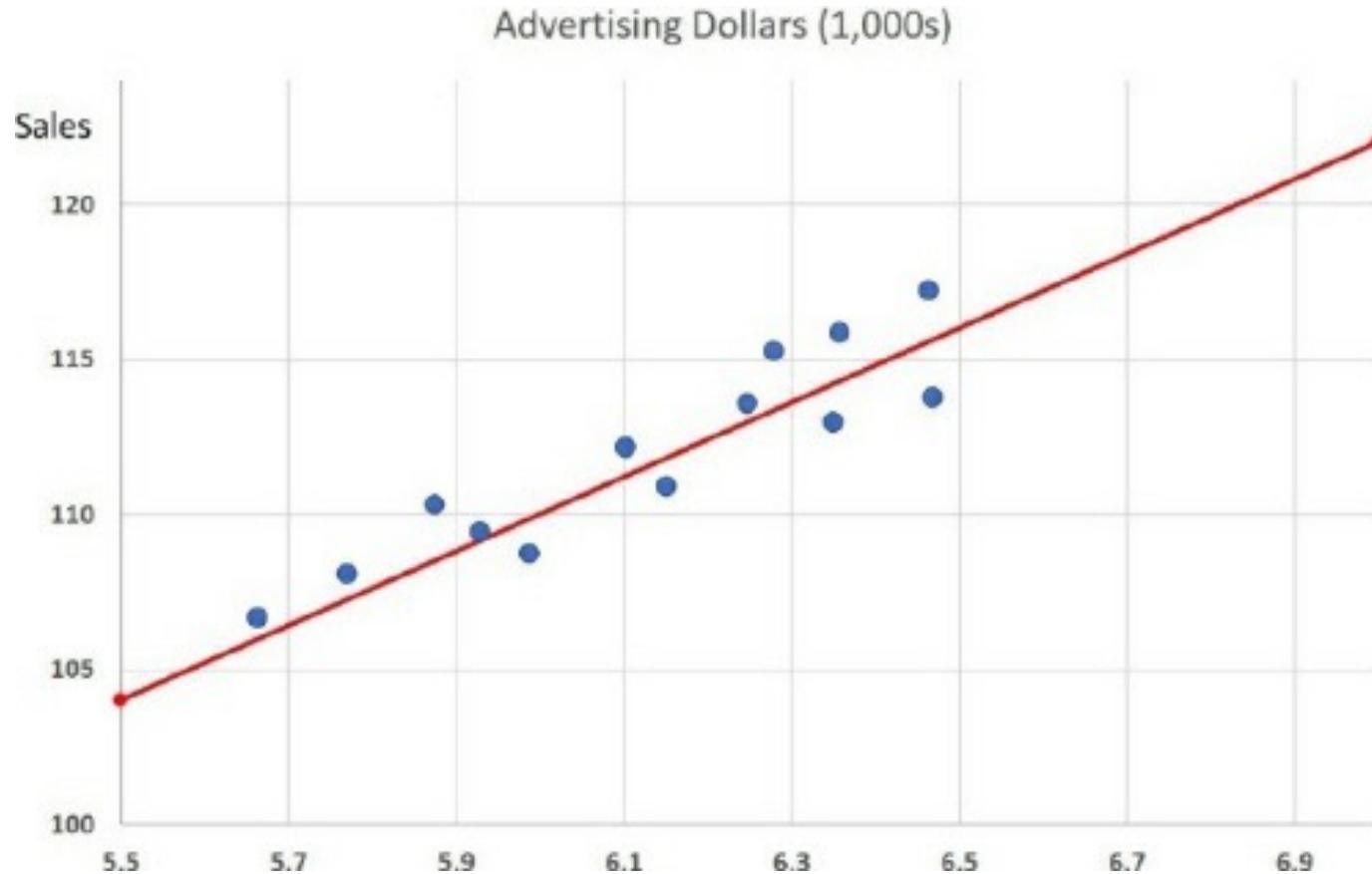
UNIVERSITÉ PARIS 1

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 ?





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

ÉCOLE DE MANAGEMENT
DE LA SORBONNE



WE FOUND THIS CORRELATION
IN THE DATA. EVERYONE
TAKE A RAZOR.





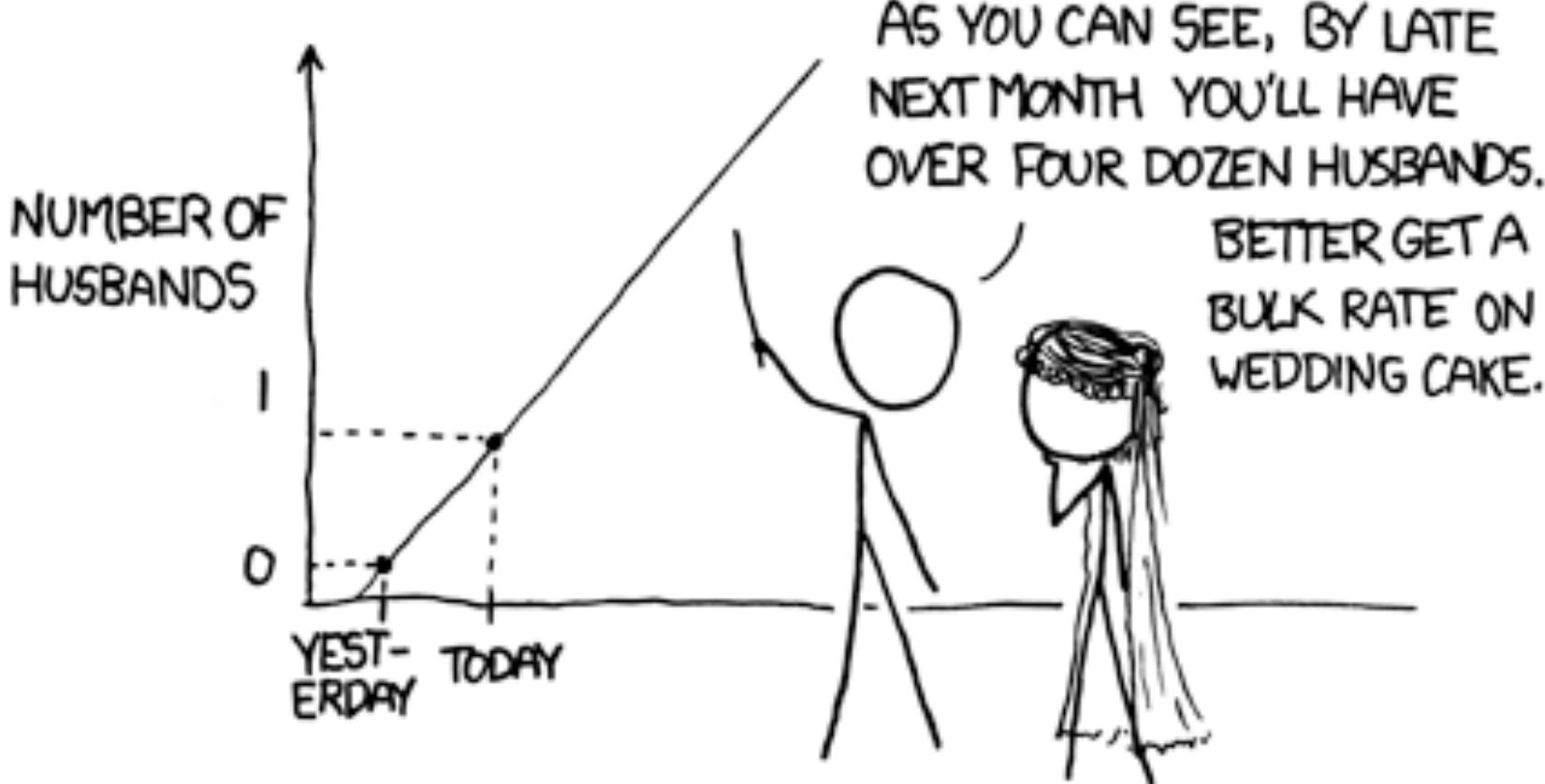
UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

ÉCOLE DE MANAGEMENT
DE LA SORBONNE



MY HOBBY: EXTRAPOLATING





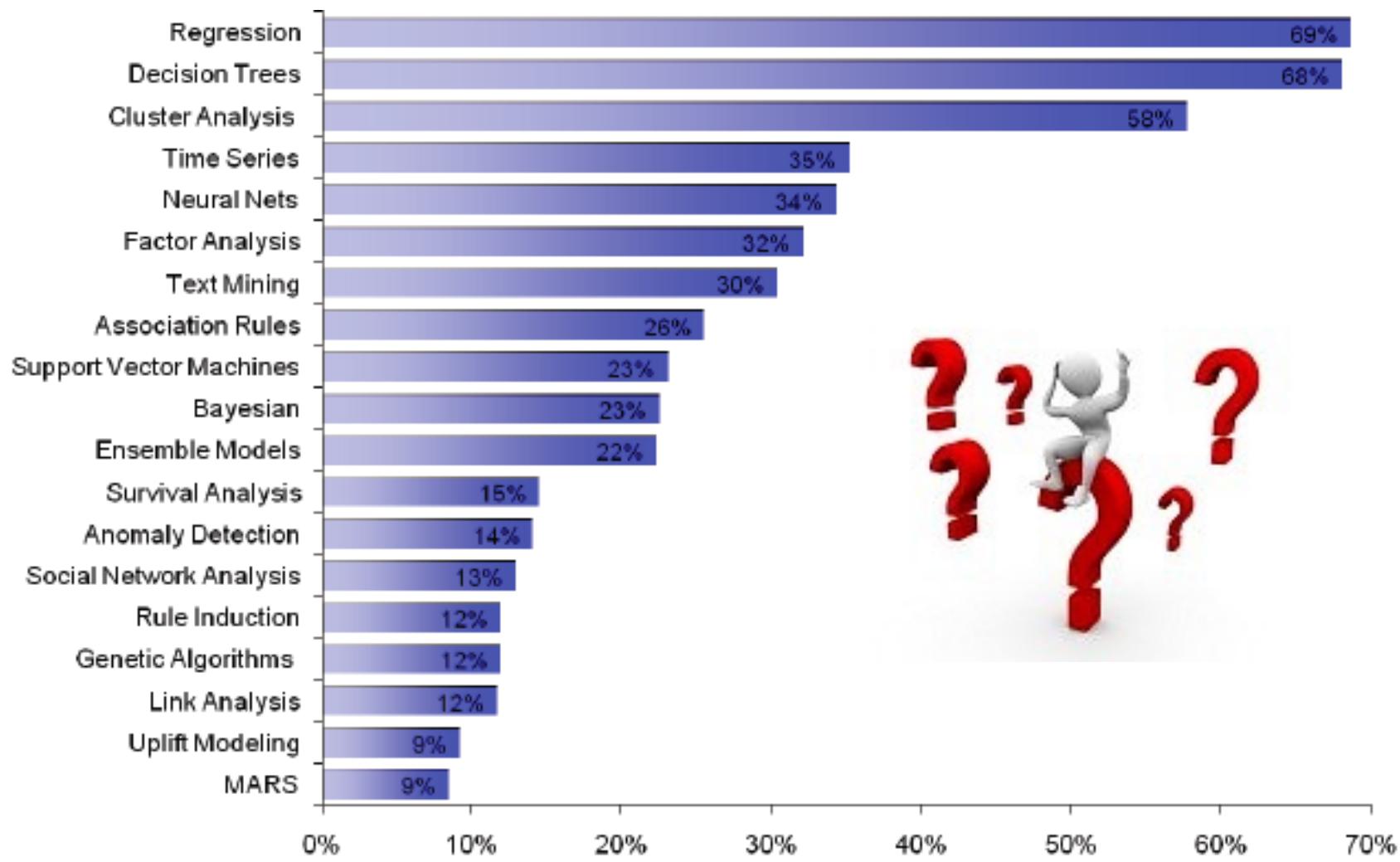


UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

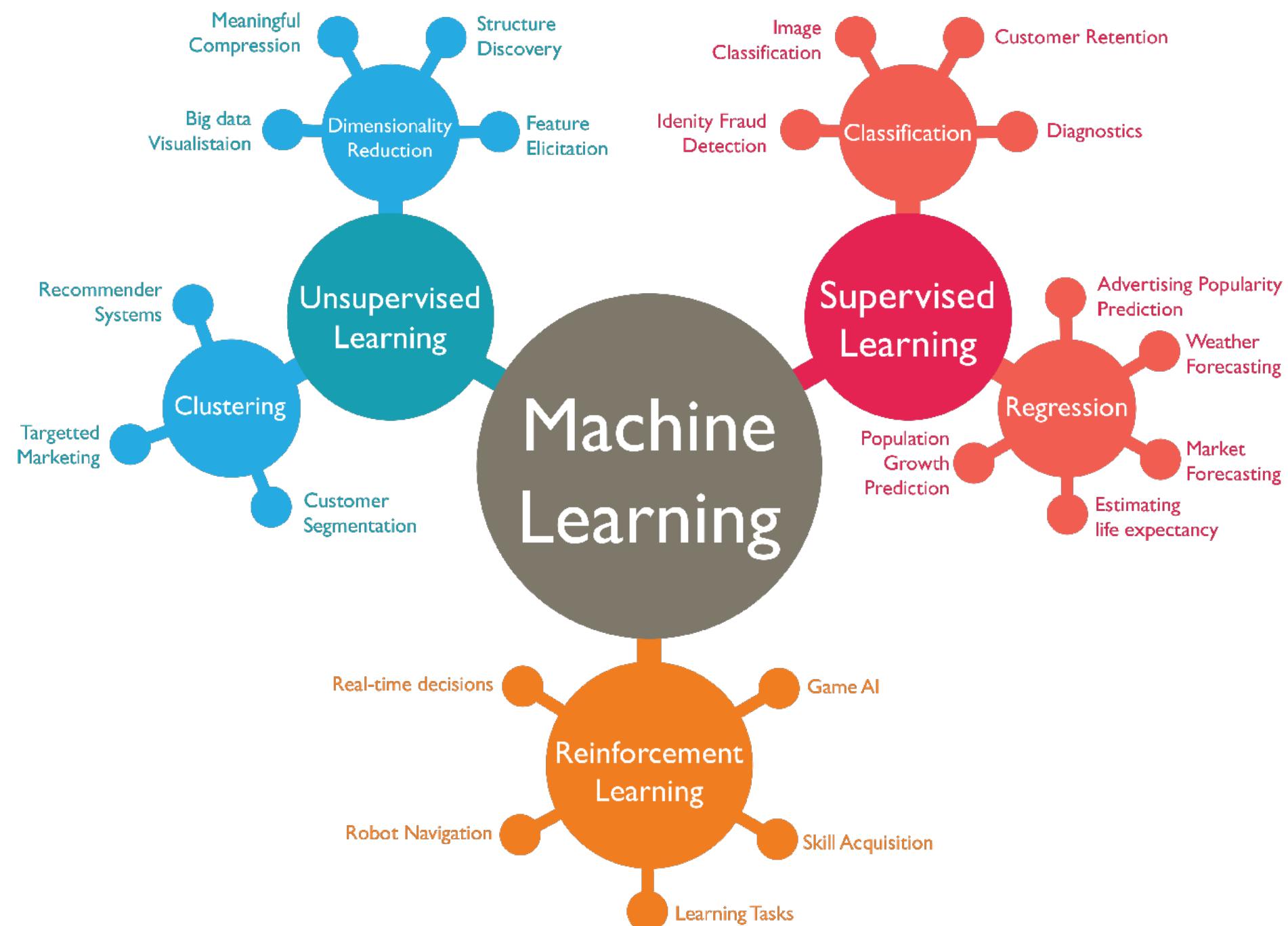
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Quel algorithme utiliser ?

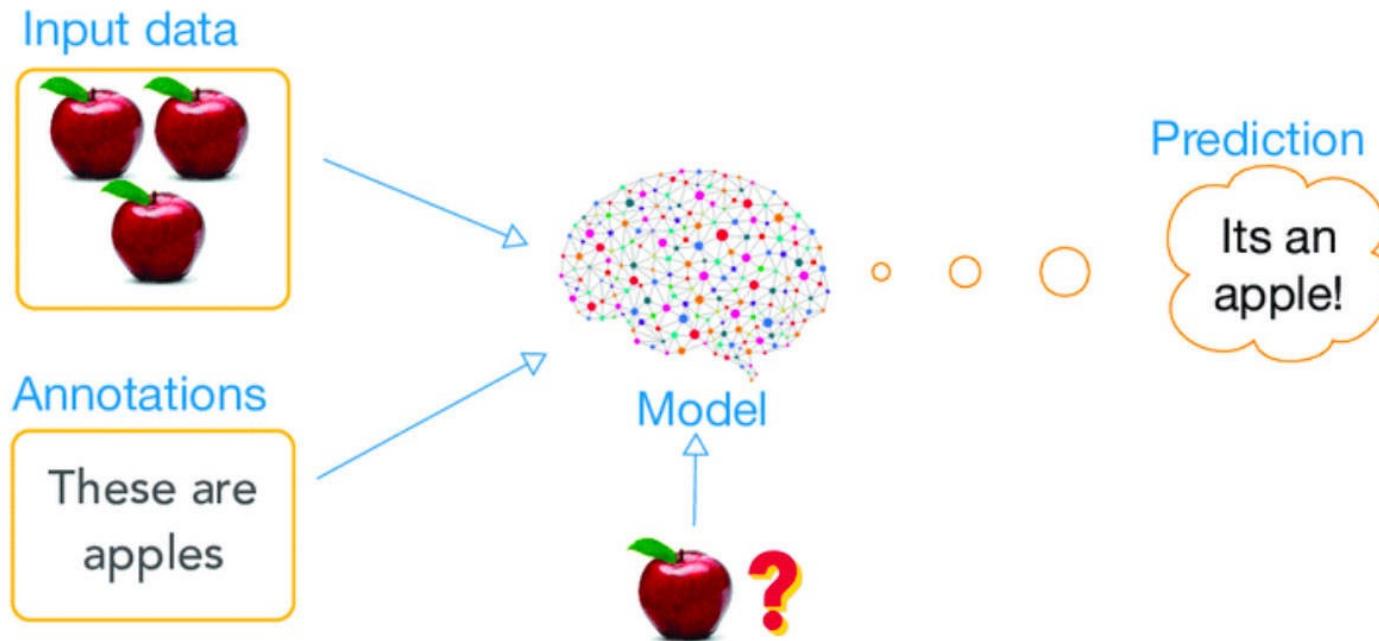


Question: What algorithms/analytic methods do you TYPICALLY use? (Select all that apply)

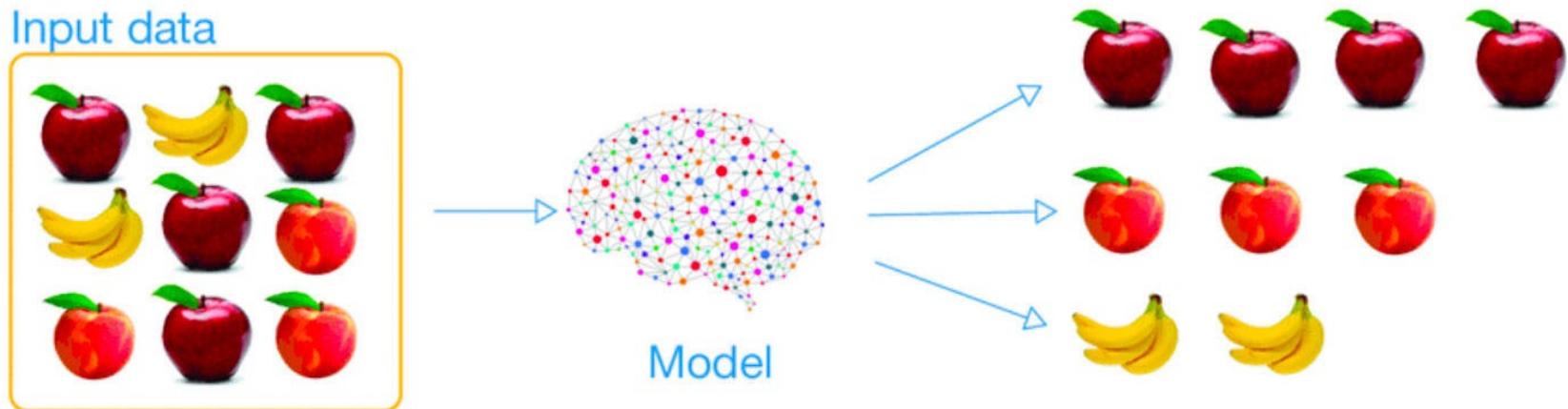
©2012 Rexter Analytics



supervised learning



unsupervised learning

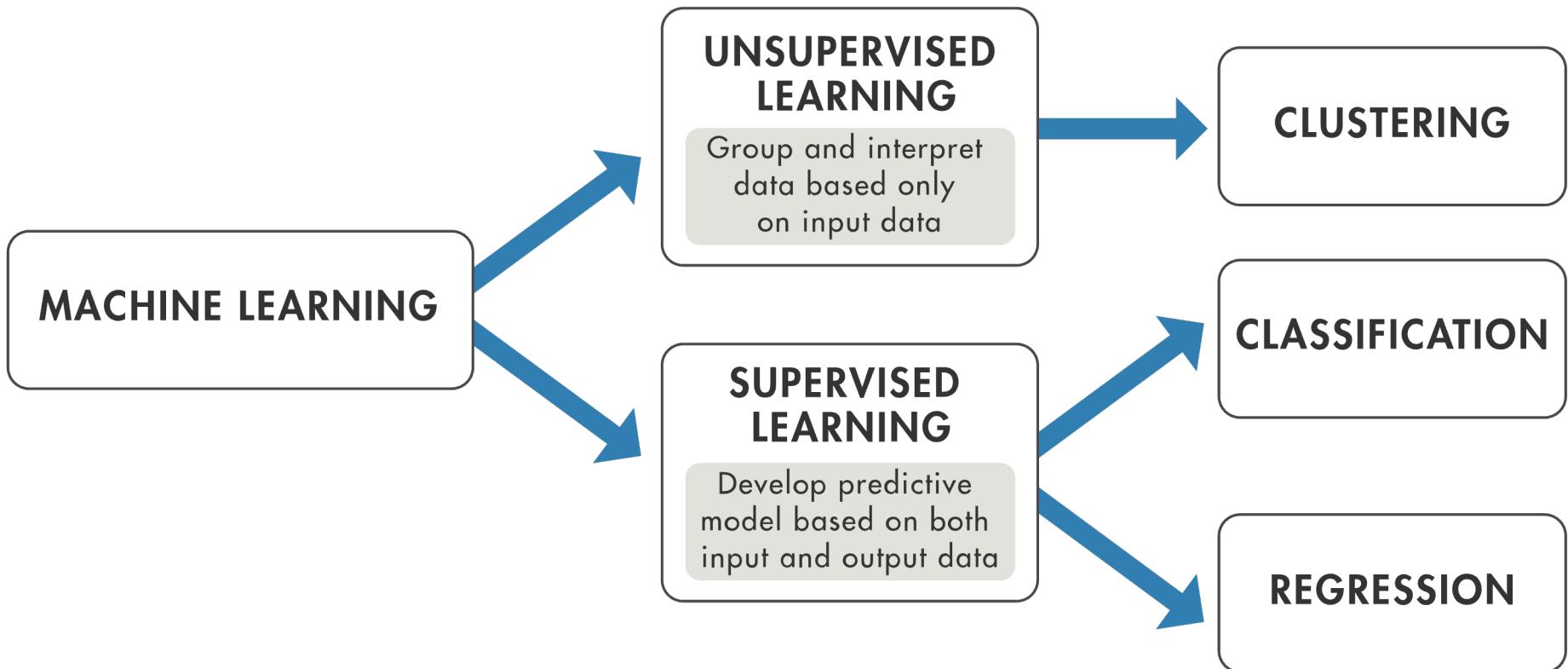




UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

ÉCOLE DE MANAGEMENT
DE LA SORBONNE

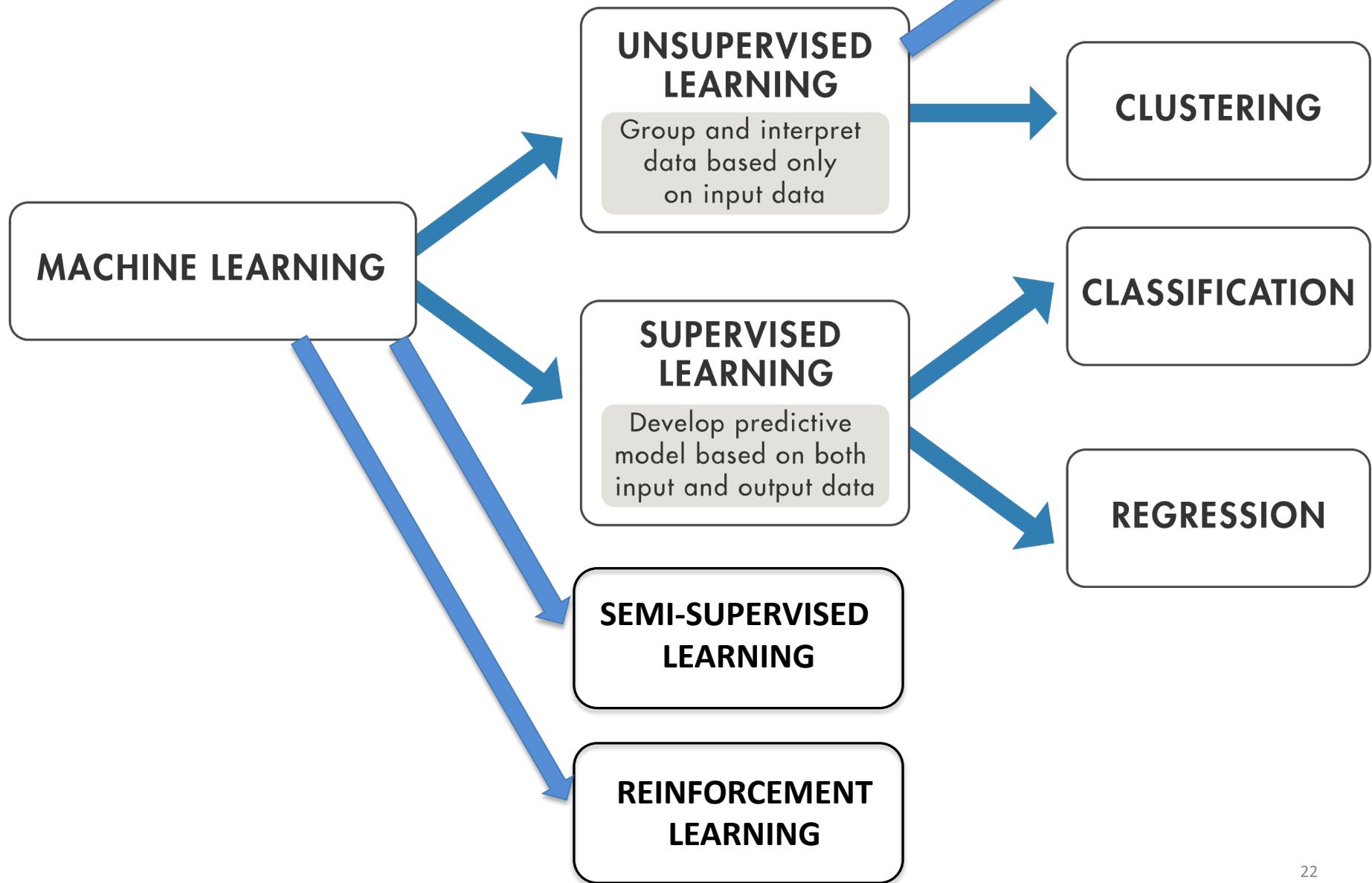




UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

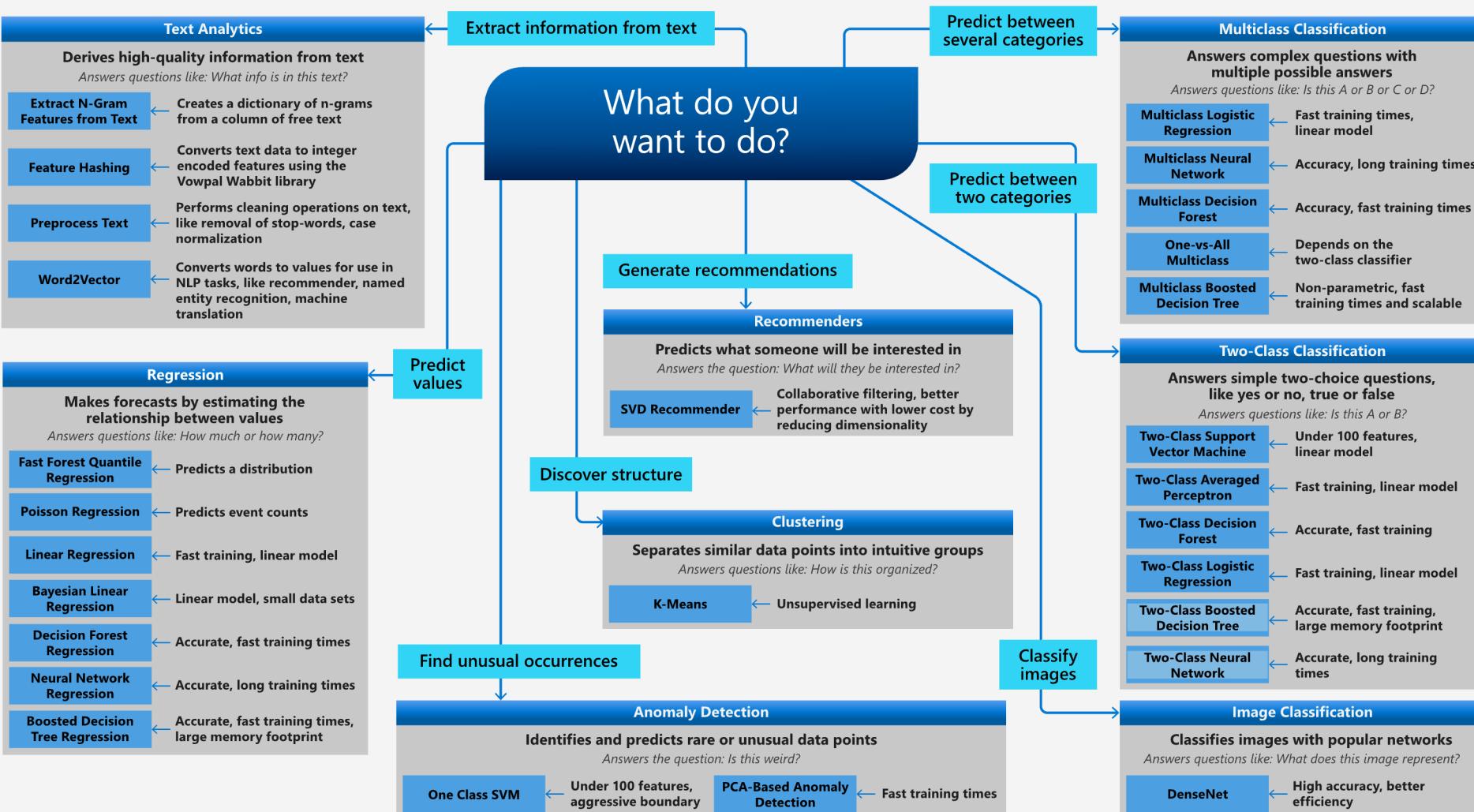
ÉCOLE DE MANAGEMENT
DE LA SORBONNE





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.



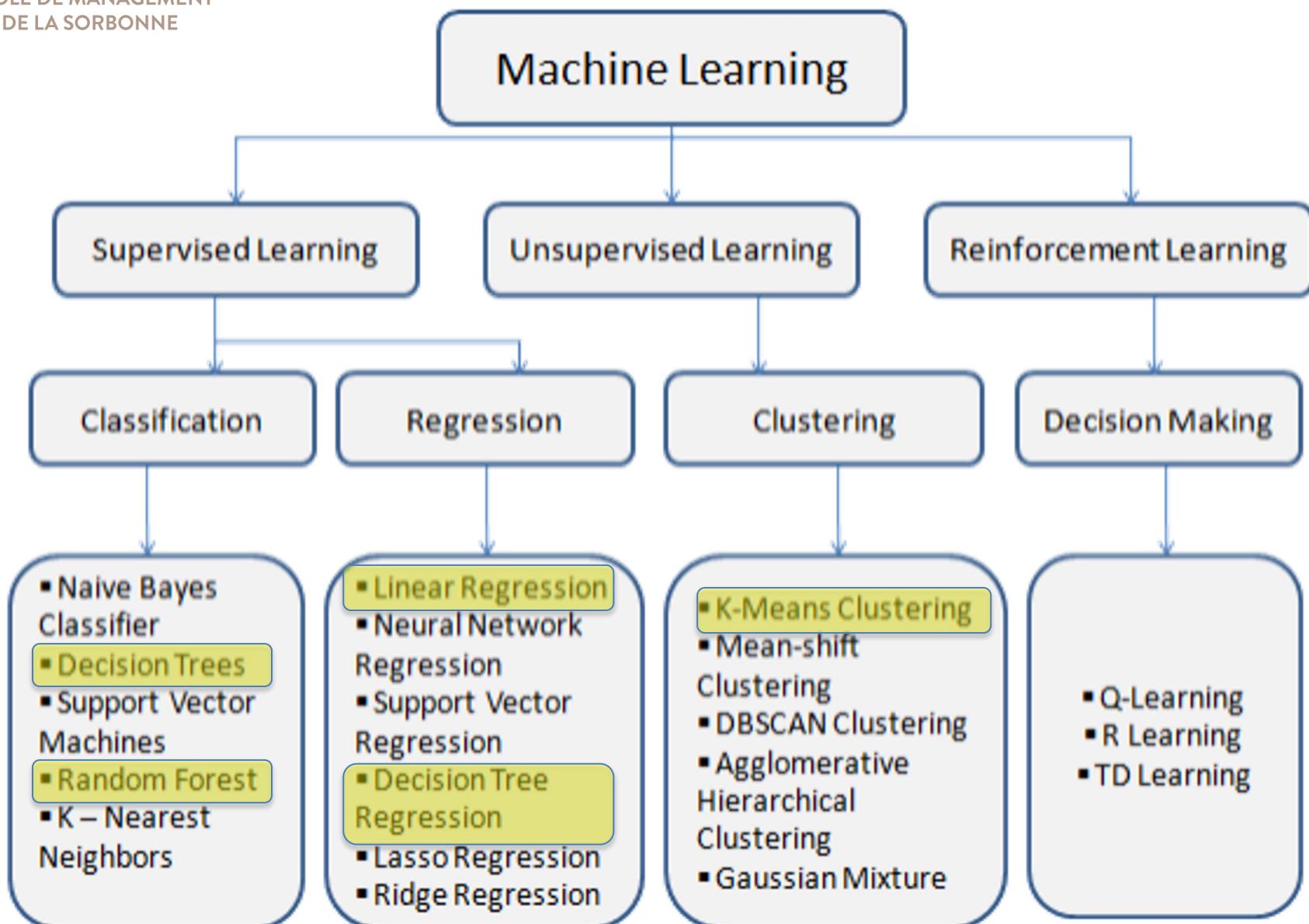


UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

ÉCOLE DE MANAGEMENT
DE LA SORBONNE

Types of Machine Learning



THE DATA SCIENCE PROCESS



Bienvenue dans l'aventure !!!