

# Aide à la décision (quelques outils de machine learning)

## BUT3 Informatique

1

[noelle.bru@univ-pau.fr](mailto:noelle.bru@univ-pau.fr)

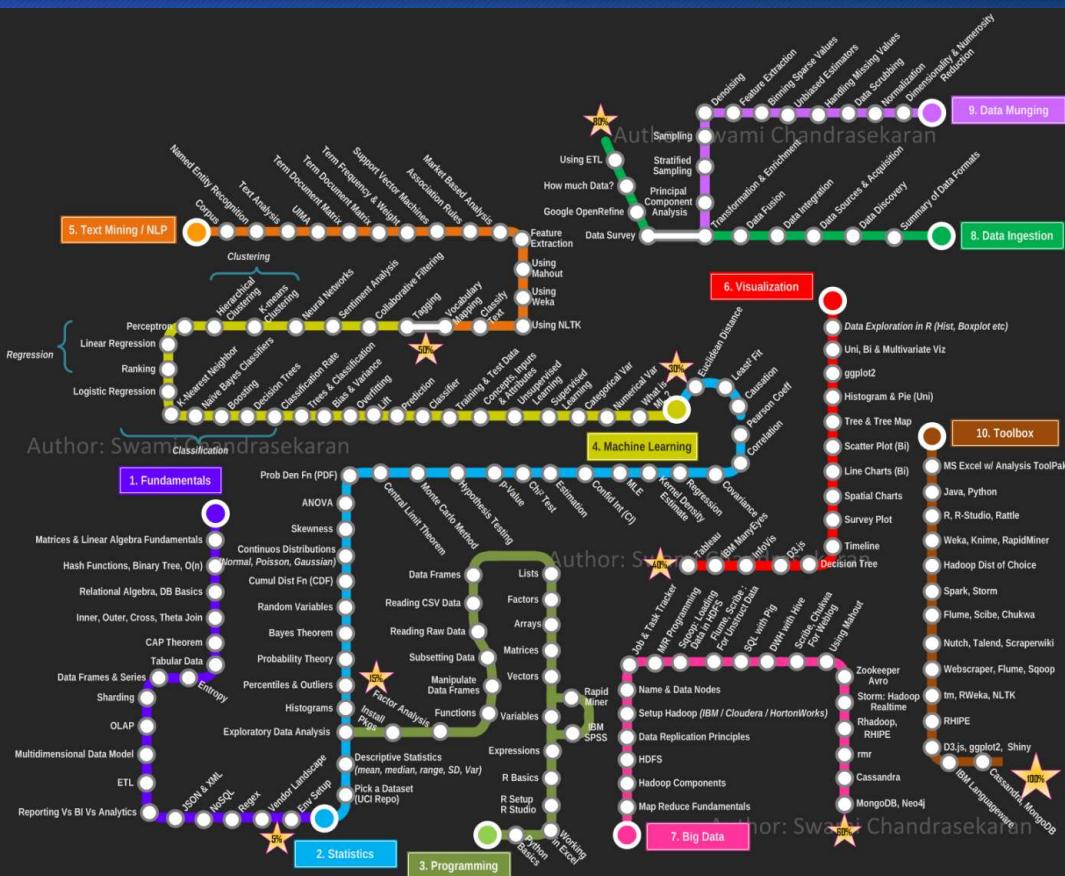


## Contexte

2

# Une compétence des « data scientists/analysts »

3



<http://nirvacana.com/thoughts/2013/07/08/becoming-a-data-scientist/>



Image Source: linkedin.com

# Un zoom...

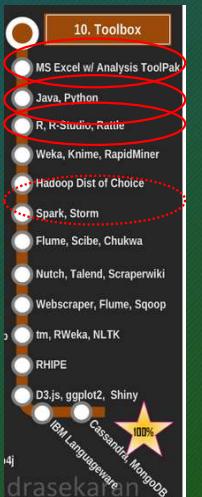
## Les méthodes



# Les outils numériques



Sipina ,  
tanagra,  
Python...



Logiciels /  
Applications/  
Langages

Gestion de la donnée

Visualisation



# Quelques méthodes... dans vos bagages...

## Méthodes descriptives

- Méthodes non supervisées :
  - Détection d'associations
  - Méthodes de partitionnement / clustering (k-means, hiérarchique)
  - ...
- Réduction de dimension
  - Méthodes factorielles
  - ...
- ...

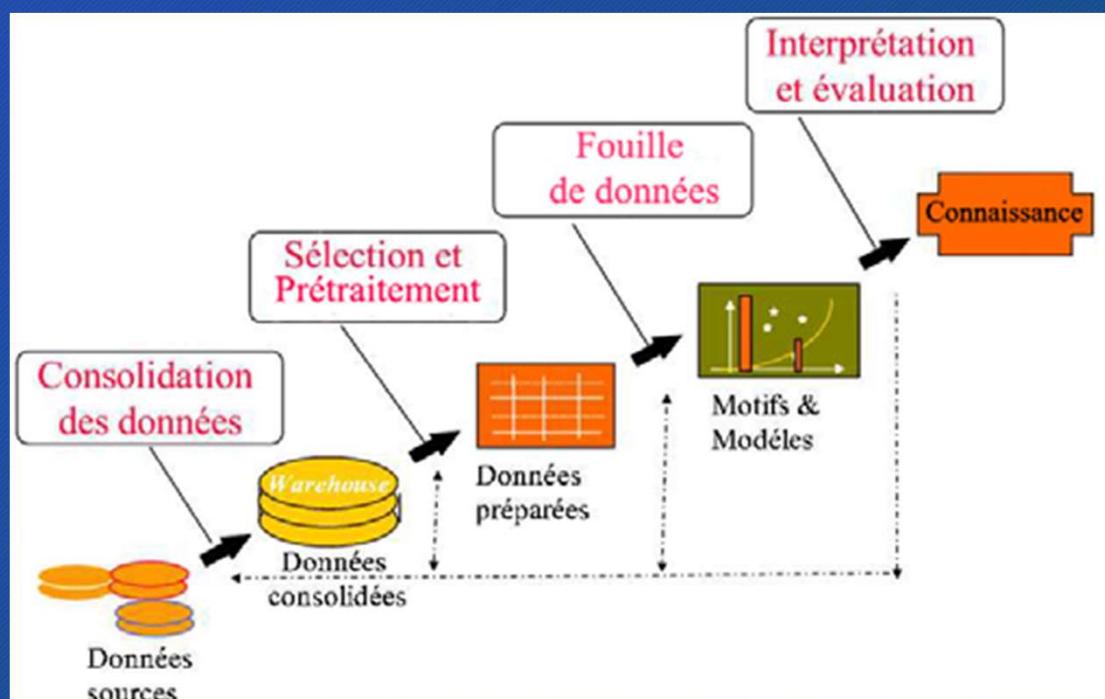
<https://datascientest.com/machine-learning-tout-savoir>

## Méthodes prédictives

- Méthodes supervisées
  - Arbre de décision / arbre de régression => Forêts aléatoires
  - Régressions linéaires et autres modèles de régression
  - ...
- Réseaux de neurones
- ...

<https://www.lemagit.fr/conseil/Machine-Learning-les-9-types-dalgorithmes-les-plus-pertinents-en-entreprise>

# La procédure Data Science



Une question concrète



Des données en général mixtes



Un ensemble de méthodes



Une recette



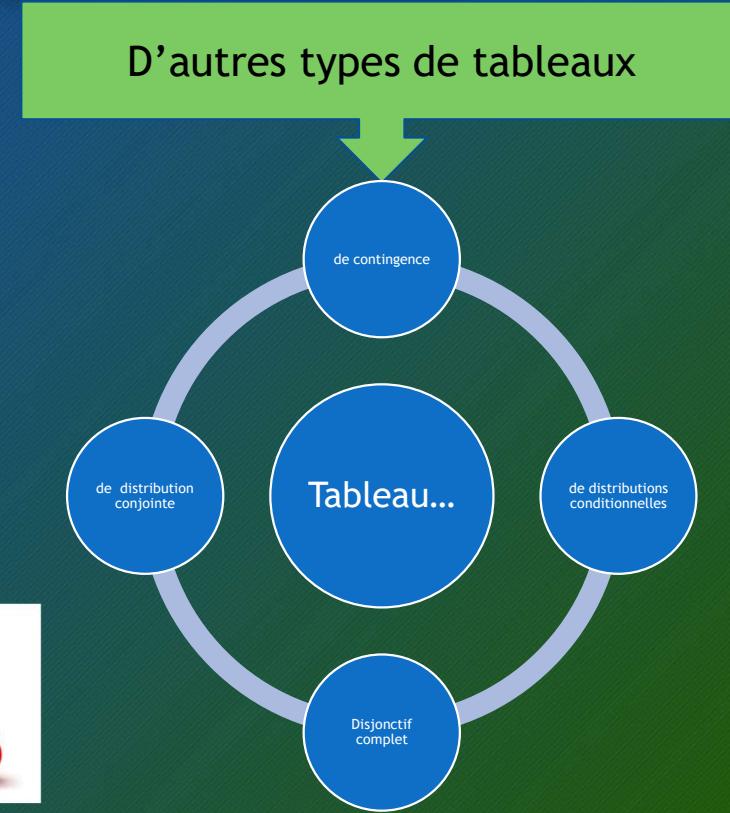
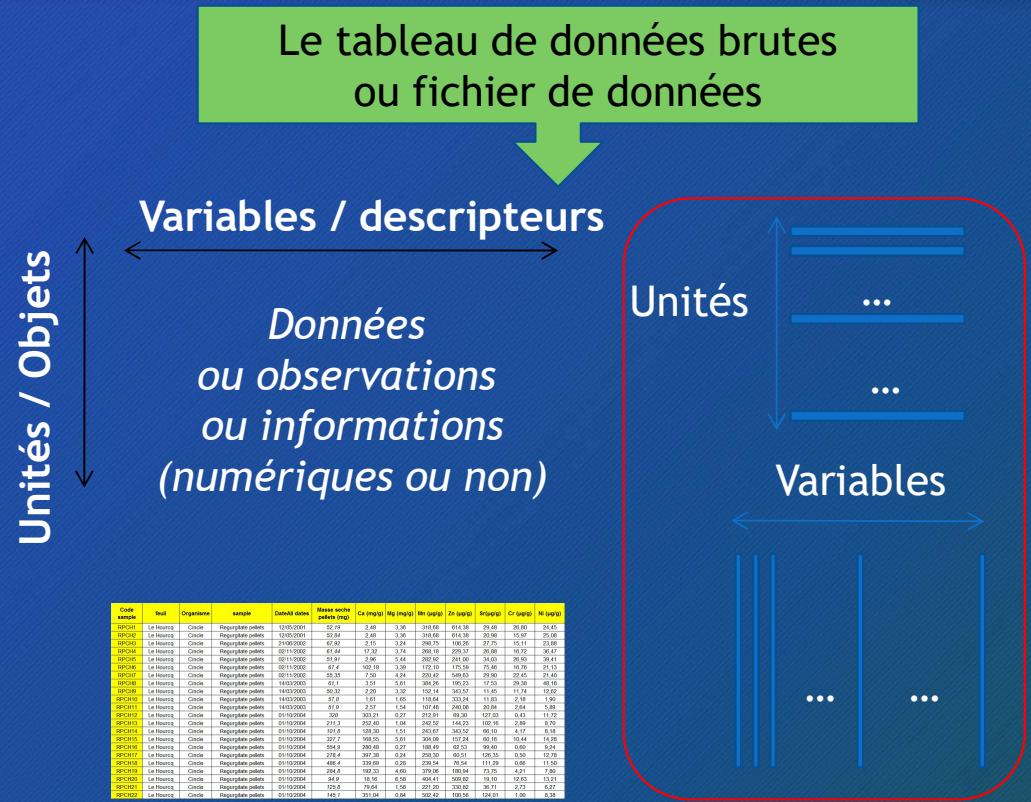
Une réponse à la question



# Au commencement... les données



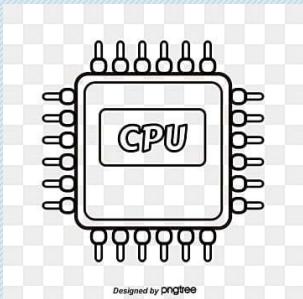
7



# Application : Relative CPU Performance Data

vendor name	Model Name	MYCT	MMIN	MMAX	CACH	CHMIN	CHMAX	PRP	ERP
burroughs	b1955	167	524	2000	8	4	15	19	23
ibm	4381-2	17	4000	16000	32	6	12	133	116
amdaHL	580-5850	23	16000	32000	64	16	32	489	381
magnuson	m80/43	50	2000	16000	24	6	16	54	88
dg	eclipse:c/350	700	384	8000	0	1	1	24	34
cdc	cyber:170/815	50	2620	10480	30	12	24	32	74
formation	f4000/300ap	800	256	8000	0	1	4	22	34
cambex	1636-10	50	1000	8000	8	3	5	36	41
ncr	v8545:ii	84	1000	2000	0	1	6	12	21
honeywell	dps:6/35	330	1000	3000	0	2	4	16	23

Phillip Ein-Dor and Jacob Feldmesser



<https://archive.ics.uci.edu/dataset/29/computer+hardware>



## Attribute Information:

1. vendor name: 30  
(adviser, amdaHL, apollo, basf, bti, burroughs, c.r.d, cambex, cdc, dec, dg, formation, four-phase, gould, honeywell, hp, ibm, ipl, magnuson, microdata, nas, ncr, nixdorf, perkin-elmer, prime, siemens, sperry, sratus, wang)
2. Model Name: many unique symbols
3. MYCT: machine cycle time in nanoseconds (integer)
4. MMIN: minimum main memory in kilobytes (integer)
5. MMAX: maximum main memory in kilobytes (integer)
6. CACH: cache memory in kilobytes (integer)
7. CHMIN: minimum channels in units (integer)
8. CHMAX: maximum channels in units (integer)
9. PRP: published relative performance (integer)
10. ERP: estimated relative performance from the original article (integer)