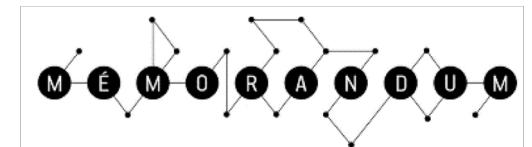


# **« Big data » et « Machine learning », en quoi le contrôle de gestion est-il concerné ?**

25 mars 2016



## Un atelier de rencontre entre finance d'entreprise et big data

### Rappel du déroulement

14h00 Introduction

14h10 Atelier : Les problématiques rencontrées en contrôle de gestion et finance d'entreprise.

14h30 Big Data et Machine Learning, de quoi s'agit-il? Exemples concrets d'applications business.

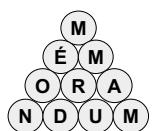
15h20 Atelier : Quels liens avec les enjeux de la finance d'entreprises? Choix de 3 sujets.

• 15h30 Pause

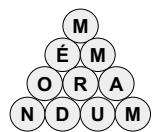
15h45 Reformulation et analyse des 3 sujets sélectionnés.

16h45 Bilan

17h00 Clôture



## Présentation des participants



**1<sup>er</sup> atelier : Quelles sont les problématiques que vous rencontrez régulièrement dans l'exercice des métiers de la finance d'entreprise ?**

## Chaine de production

## **Acquisition : des données jamais objectives**

- de très forte variations culturelles / locales biaissent la collecte à la source (qu'est ce qu'un incident ?)

## Agrégation : un biais systématique

- Variations locales des périmètres d'agrégation (toujours une marge de liberté)
  - normes locales, taux de changes et couvertures
  - Contraintes opérationnelles (cut off, « black out »)

## Accès à la donnée : 2 barrières à lever

- SI : multiplicité des systèmes
  - Culture : Chaque direction se vit comme propriétaire de ses données, contrôle de gestion cloisonné en finance

## Interprétation : le flou

- Trop de données : que faut-il analyser
  - Quels KPI pertinents, vis-à-vis de quoi ?

## Mission

## **Le Réglementaire des suffit à lui même**

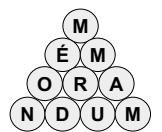
- exigences externe de « high frequency reporting »
  - directions opérationnelles sur la défensive



→ production et à justification d'indicateurs de synthèse avec peu ou pas d'impact opérationnel

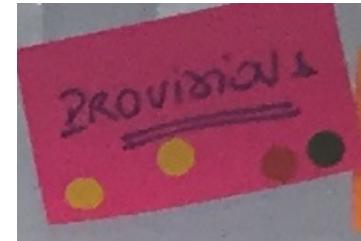
**Enjeu est dans le diagnostic opérationnel interne :  
médecin prescripteur**

## Quelques définitions métier (cf guillaume)



## 2<sup>e</sup> atelier : cibles prioritaires du big data

### → 1) provisions pour créance client



Les provisions comptables ne peuvent être le résultat d'une analyse statistique

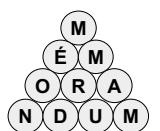
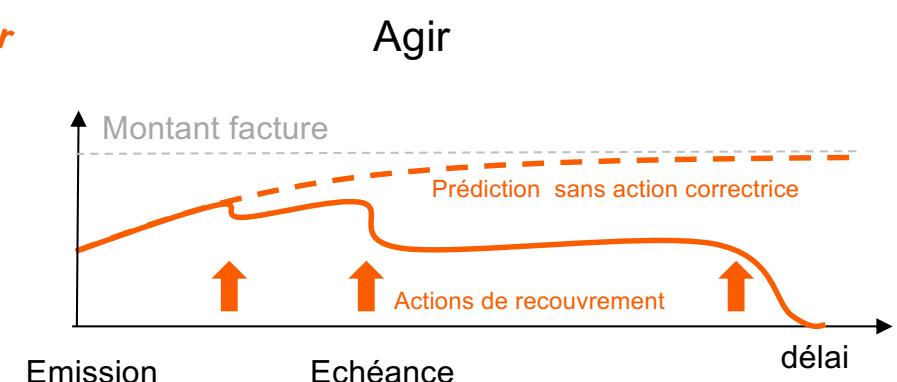
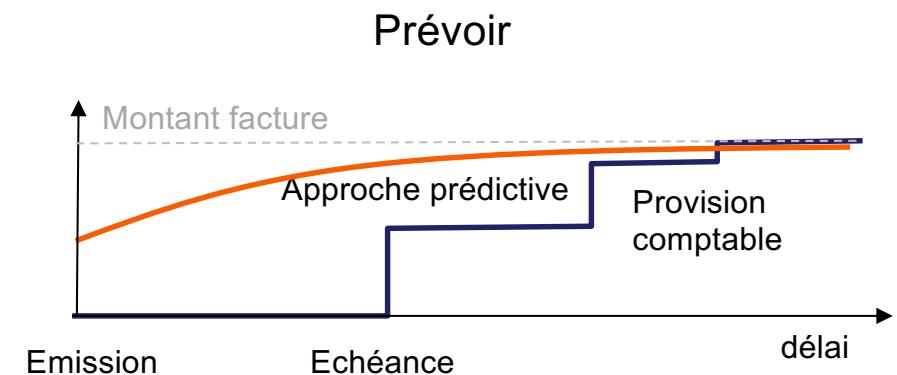
*Constater*

- Cible binaire (facture sera payée ou non) plus que de trésorerie (date de paiement)
  - Application de règles comptables strictes (0 jusqu'échéance, puis augmentation selon niveau de retard)
- Enjeu strictement comptable

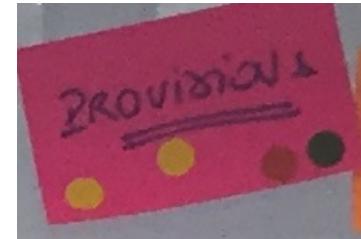
Une approche prédictive a un objectif d'annoncer dès le départ une prédition la plus juste possible de la provision finale, afin :

- D'anticiper le risque dès la vente (montant / mode et condition de règlement, ...)
- D'orienter le cas échéant les moyens de recouvrement
  - Canal (appel / courrier)
  - Message
    - levier : paiement immédiat / délai / échelonnement / ...
    - Arguments

→ Enjeu opérationnel d'allocation de ressources recouvrement



## 2<sup>e</sup> atelier : cibles prioritaires du big data → 2) autres provisions



### dépréciation de stock

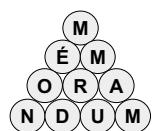
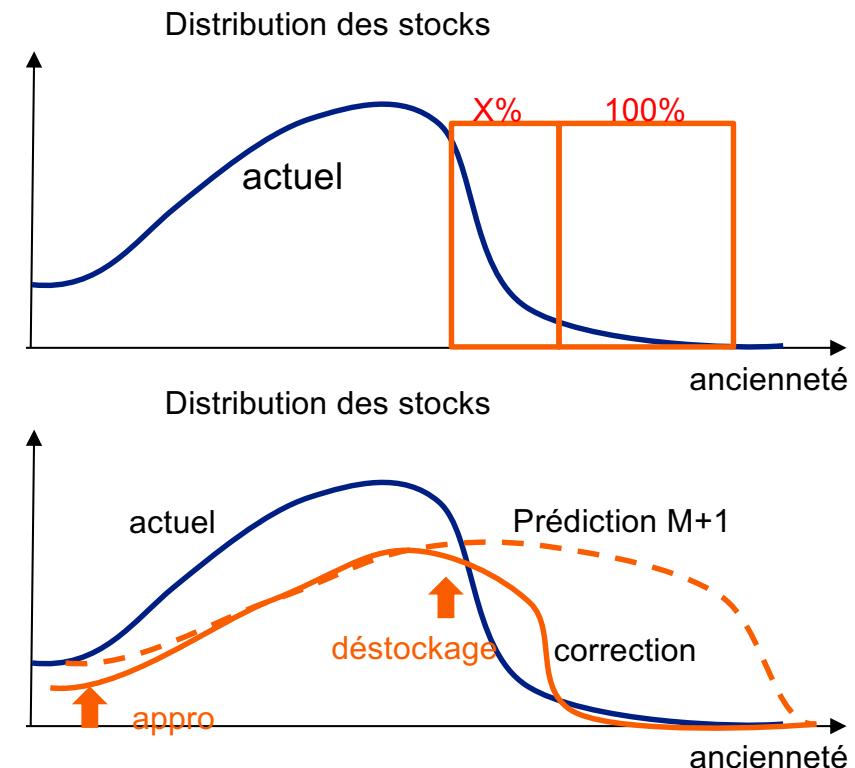
Les provisions comptables constatent une ancienneté des stocks et appliquent des règles de dépréciation

Une approche prédictive a un objectif d'anticiper la structure de dépréciation du stock suffisamment tôt pour lancer des actions correctrices

- Déstockage
  - Ajustement des appro
  - ...
- Enjeu opérationnel d'allocation de ressources recouvrement

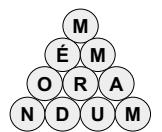
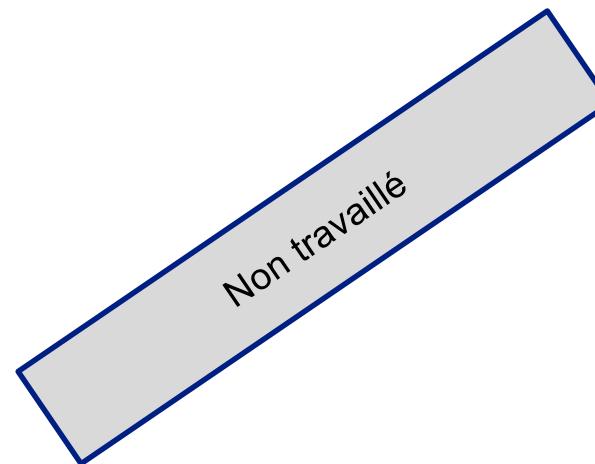
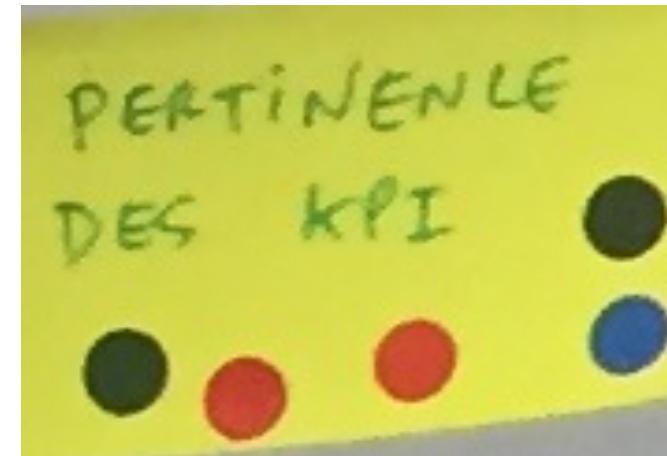
La même approche peut être déclinée sur d'autres provisions

- Indemnités de retraites, ..



## 2è atelier : cibles prioritaires du big data → 3) identifier les bons KPI

La sélection de KPI pertinent est à la base du machine learning ...



## Annexe : présentation faite en séance sur le big data

14h00 Introduction

14h10 Atelier : Les problématiques rencontrées en contrôle de gestion et finance d'entreprise.

- 14h30 **Big Data et Machine Learning, de quoi s'agit-il? Exemples concrets d'applications.**

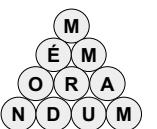
15h20 Atelier : Quels liens avec les enjeux de la finance d'entreprises? Choix de 3 sujets.

- 15h30 Pause

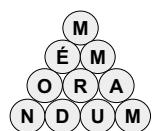
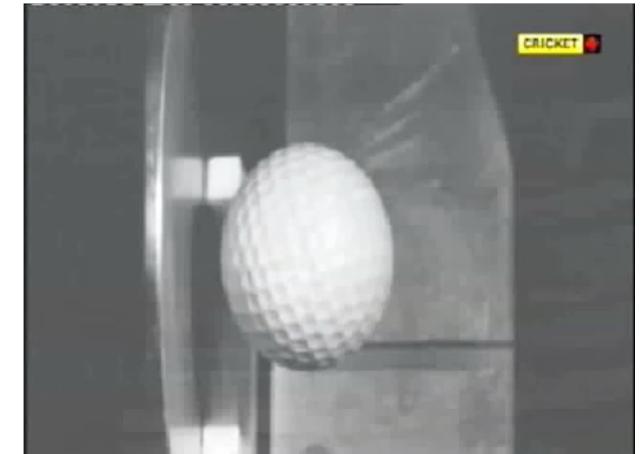
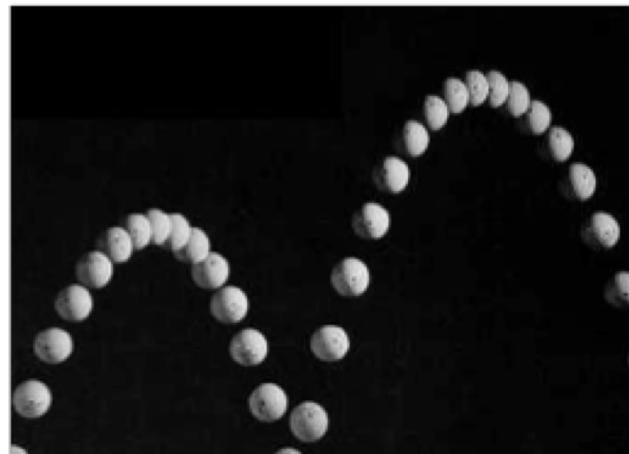
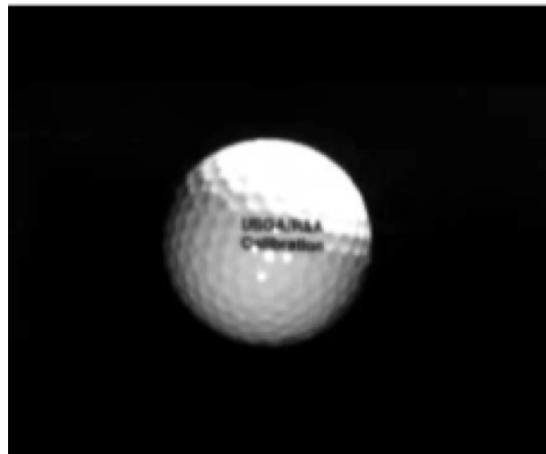
15h45 Reformulation et analyse des 3 sujets sélectionnés.

16h45 Bilan

17h00 Clôture



**Connaissons-nous le monde ?**



## Une rupture en cours

*Les technologies de modélisation ont occupé financièrement l'innovation historique faible pour ils se les fond concept que la n bien au marketing chain.*



Aremus & Associés  
CAPITAL-INNOVATION

**80% des dirigeants seront les figurants de demain.**

Et vous ?

**Oui. Votre business est déjà concerné.**

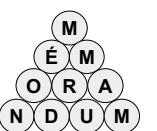
La disruption sera vecteur de compétitivité pour les entreprises prêtes à se réinventer. A condition de faire vite et de définir une vision claire, sous peine de décroître puis sombrer.

“ l'algorithme intime du temps réel : la disruption s'apprend. Heureusement, mais plus profondément encore d'entrer dans une transformation de l'expérience client sur des produits et services qui peuvent désormais bénéficier d'une évolution constante.”

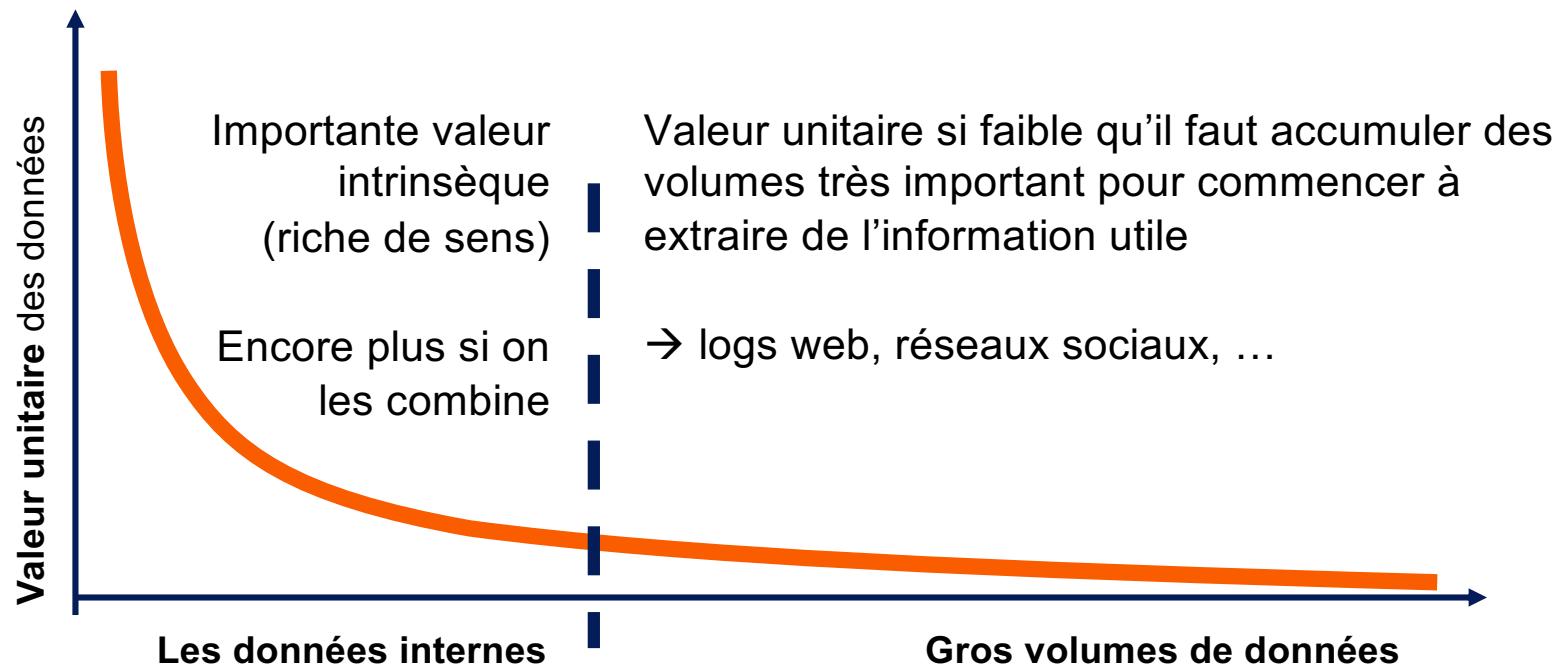
*Au fond, les modèles et algorithmes aménagent la fois des et en pour start-up. logies ement nieur, es et allant ussage nées, utili- ille et*

**bpifrance**

**Modèles, données et algorithmes**  
Les nouvelles frontières du numérique  
Par Etienne de Rocquigny (Aremus & Associés)



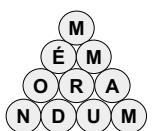
## Quelles données



Les projets de valorisation des données focalisent souvent sur les sources de gros volumes (justifiant l'infrastructure) au détriment des données à forte valeur intrinsèques, déjà disponibles et accessibles

Les données internes sont à la fois

- Concentrées et accessibles : dans l'entreprise, dans le SI
- Éclatées au sein de l'organisation



## Chacune de vos données embarque deux leviers de valeur

Une donnée parmi d'autres :  
**« Le client X a contacté le service client »**

### L'information pour elle-même

Déclencher l'action suivante :

- Répondre au client

Mettre à jour la rémunération variable du conseiller,

Alimenter les ~~rapports~~

→ **Domaine connu**

La donnée n'a été initialement produite que pour cela. Après usage cette donnée est aujourd'hui un déchet

### Le contexte d'autres événements

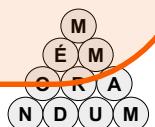
**Reconstituer le contexte** d'autres événements de l'entreprise, tous interdépendants à des degrés divers :

Une vente s'est faite entre un client donné et un produit particulier, dans un magasin spécifique, avec un vendeur unique, un certain jour de la semaine

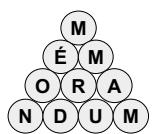
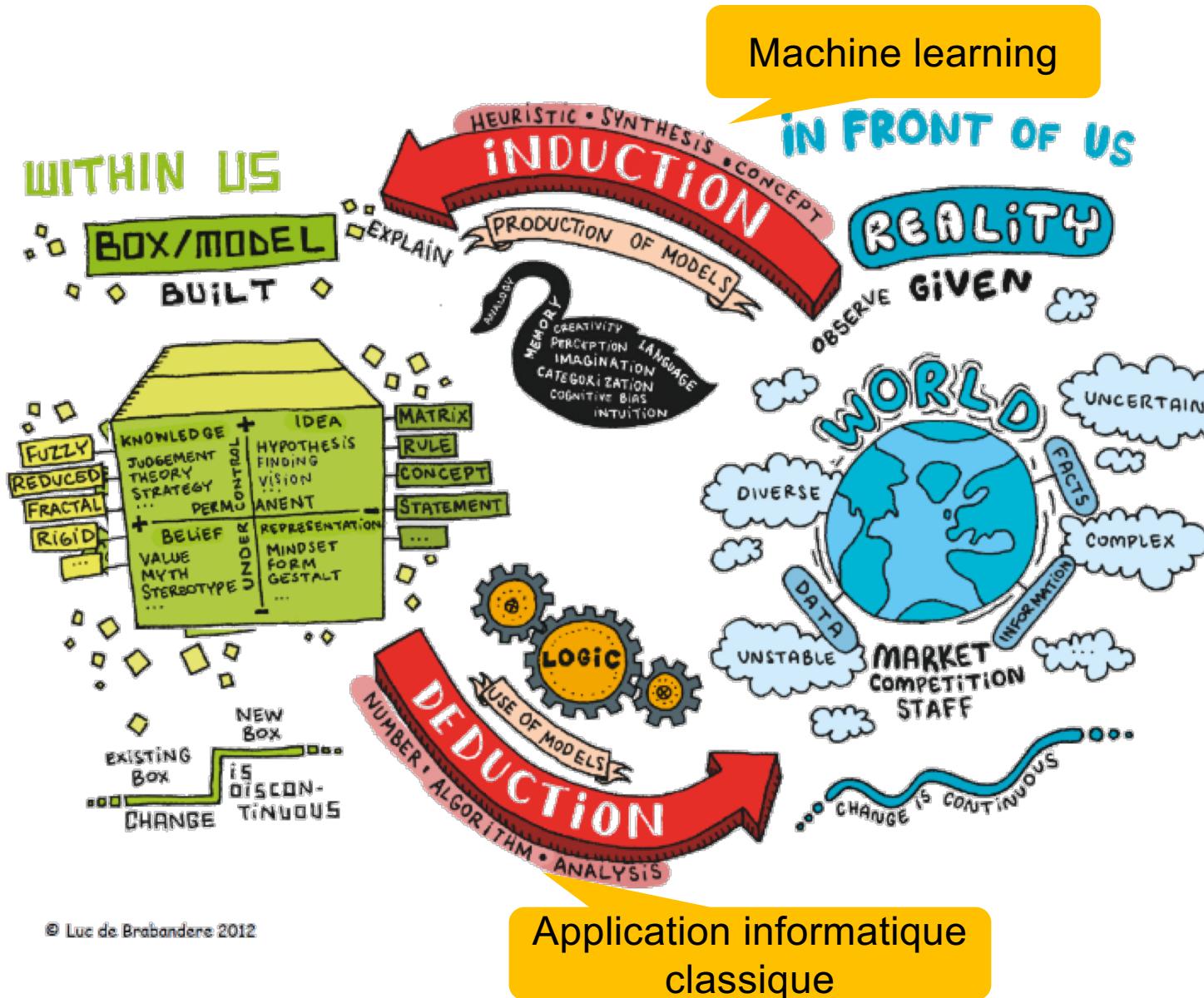
Le client s'était manifesté 3 jours avant pour dire xxx

Un exemple de préoccupation

Une donnée sortie de son contexte qui éclaire le sujet



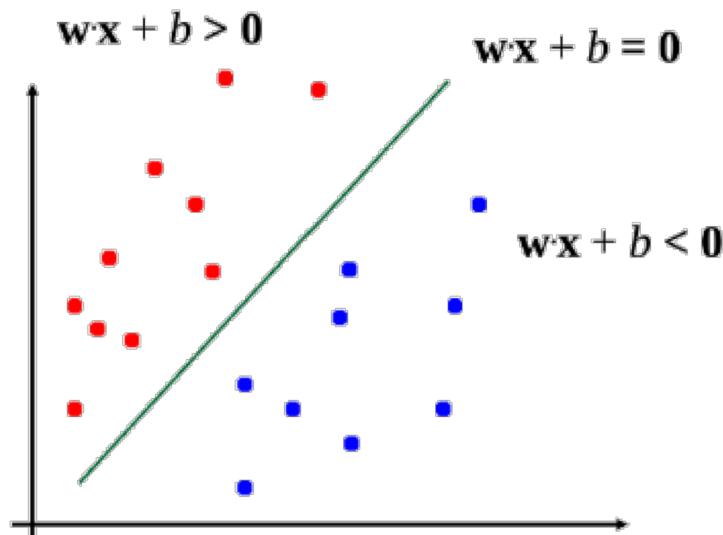
## Datascience : ni plus ni moins ce que fait votre cerveau tous les jours



## Apprentissage automatique : deux grandes familles

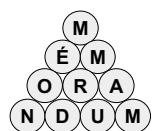
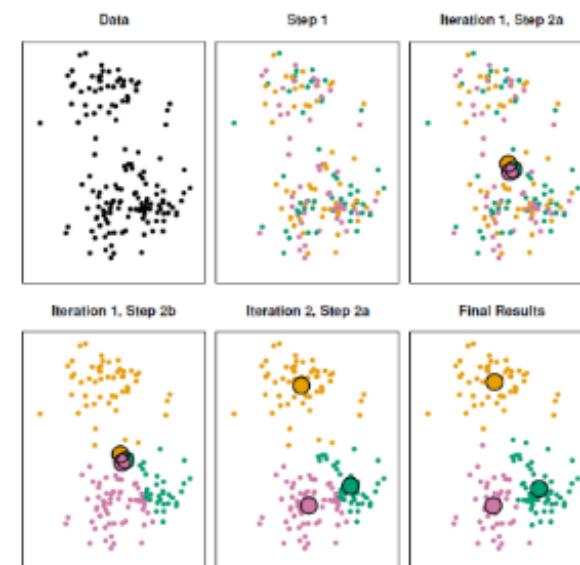
### L'apprentissage supervisé :

La machine apprend à partir d'une série d'exemples pour lesquels le résultat est connu (base d'entraînement).



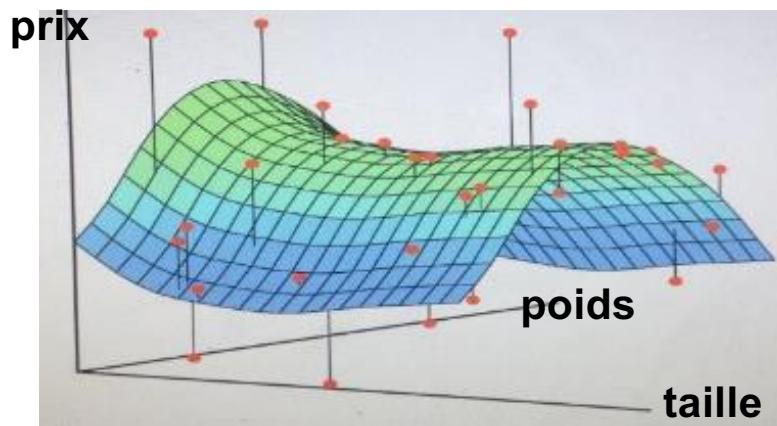
### L'apprentissage non supervisé :

La machine apprend sans exemples.

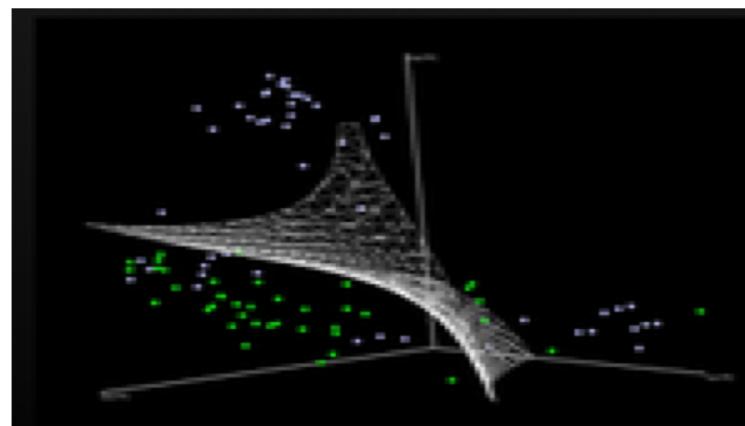


## Apprentissage supervisé : deux grandes familles de problèmes

**Quantitatif : régression,**



**Qualitatif : classification**



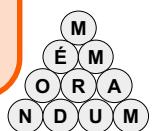
**Objectif :** Réunir les points d' un même hyperplan      Séparer les observations le plus proprement  
(droite si une feature, plan si 2 features, ... possible)

**Résultat :** Une valeur numérique

Exemple : « Apprendre » à la machine à trouver le prix d'une pierre précieuse en fonction du poids et de la taille, à partir d'une série d'exemples de transactions sur un marché .

Une classe

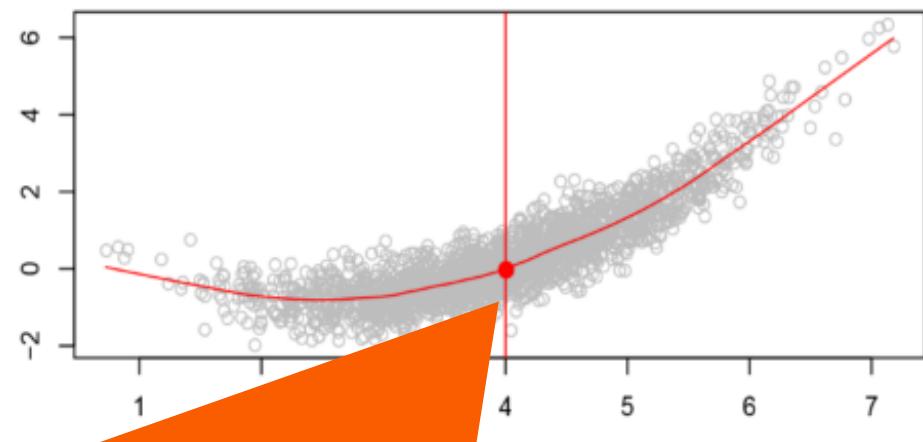
Exemple : « Apprendre » la machine à reconnaître les spam des email non spam, ou bien les pommes des poires d'une série d'images.



## Le problème (1/2)

Idéalement

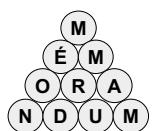
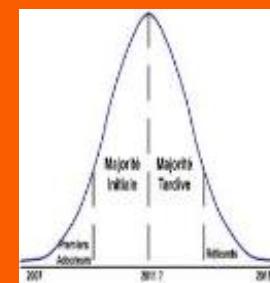
- Etre omniscient et avoir toutes les observations possibles
- Pour chaque valeur possible des features disponibles : prendre la moyenne des observations (espérance)



La dispersion autour de cette valeur moyenne peut être lié à plusieurs facteurs

Principal levier  
big data

- Il manque des facteurs explicatifs → toujours
- Il y a des erreurs de mesure → toujours
- Il y a du vrai hasard → là c'est de la philo



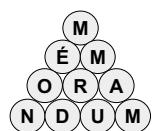
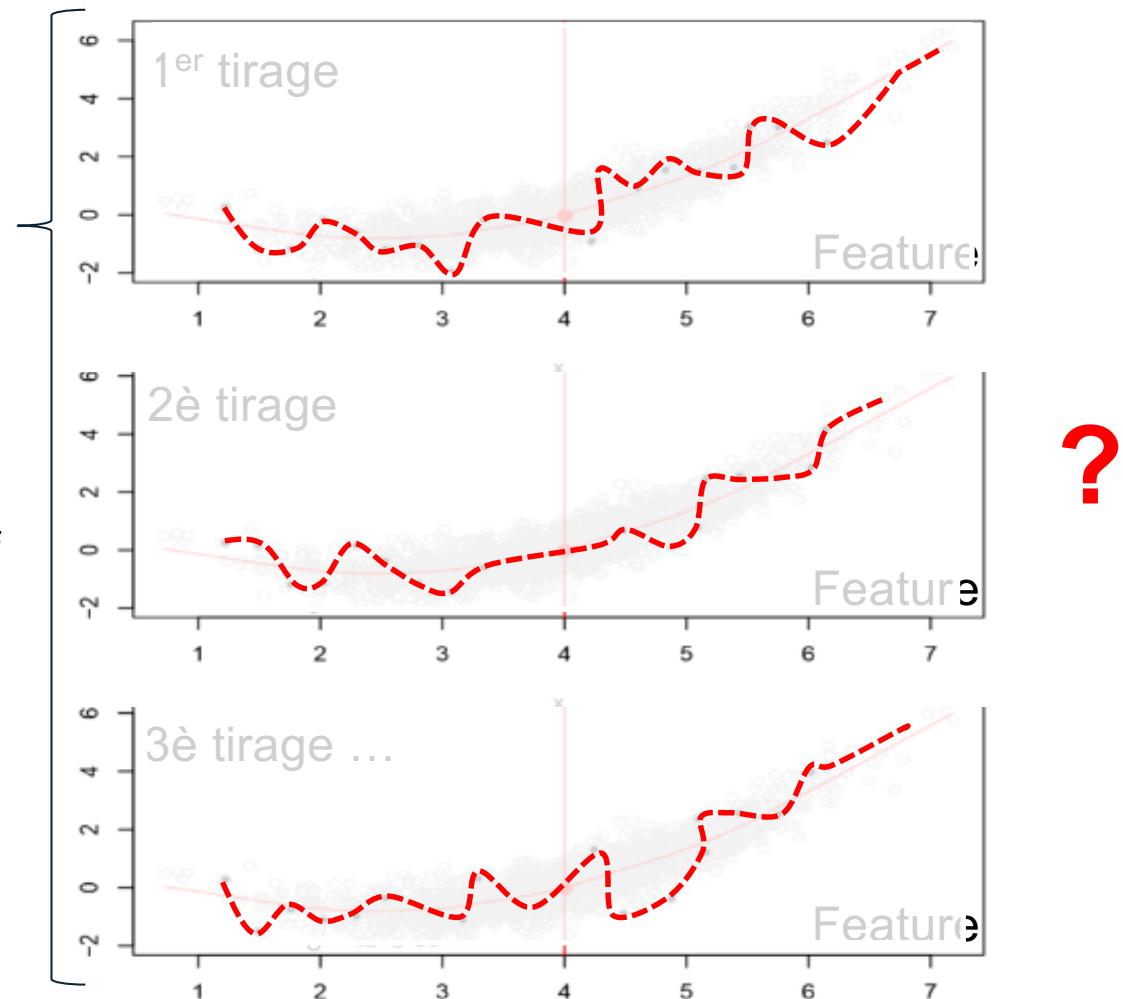
## Le problème (2/2)

**Hélas:**

- Vous ne disposez que d'un jeu de données partiel et si vous renouvez les mesures vous aurez chaque fois un autre jeu d'observation

**Hélas (bis)**

- Vous avez beaucoup d'observations.. mais encore plus de features pour chaque observation : vos êtes atteint par la malédiction de la dimension (« curse of dimensionality ») :



## Une démarche pleine de bon sens 1/3

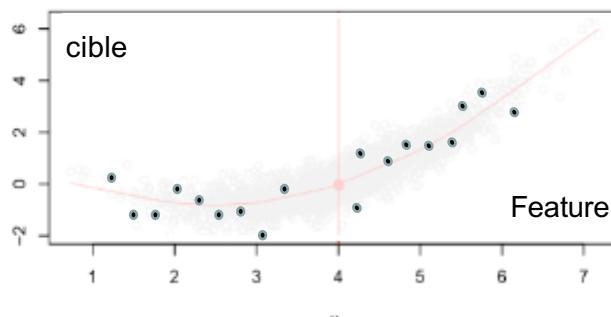
Pour bien prévoir le futur nous pouvons **simplifier le passé**

Equivalent

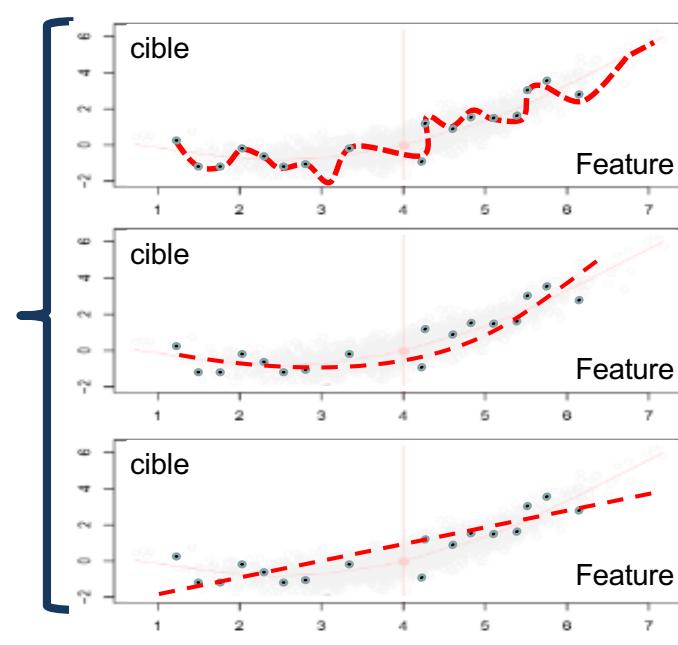
Distinguer

- le **signal** : « vraie » information apportée par les features disponibles
- du **bruit** : effet des informations (features) qui nous manquent

Données initiales



Modélisation induite



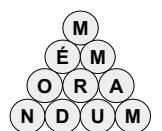
### Complexe

- Parfaite description du passé
- Faible pouvoir prédictif
- « overfitting »

Un juste milieu ?

### Simple

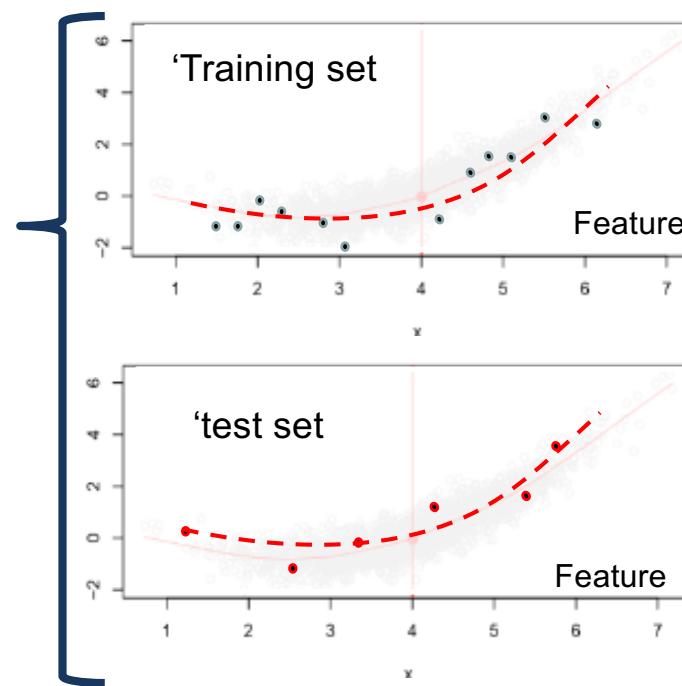
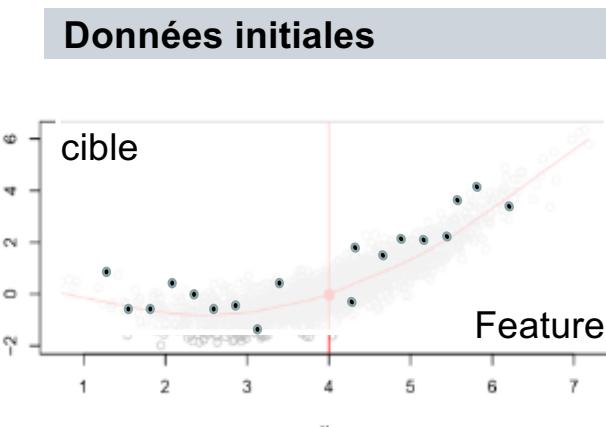
- Grossière description du passé
- Faible pouvoir prédictif



## Une démarche pleine de bon sens 2/3

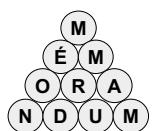
2<sup>e</sup> astuce : appliquer l'adage qui dit qu'on ne peut être juge et parti (séparation jeu d'apprentissage et jeu de test)

- L'évaluation de l'erreur d'interpolation des données connues n'est visiblement pas la métrique pertinente (sinon on va systématiquement pencher du côté « overfitting »)
- Solution « on ne peut pas être juge et partie »: les données connues sont réparties en deux lots
  - Un lot d'apprentissage
  - Un lot d'évaluation



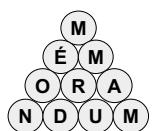
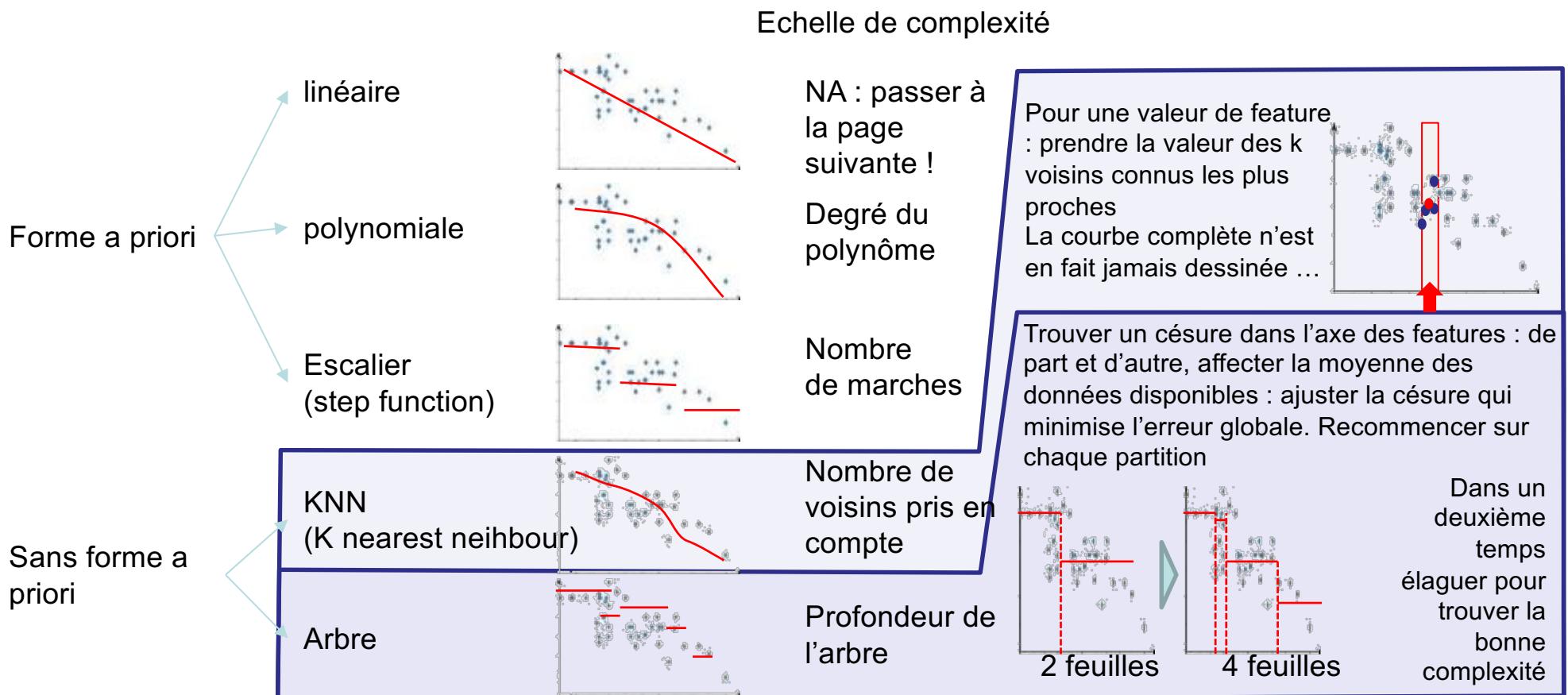
Construction d'un modèle pur un niveau de complexité donné

Evaluation : le modèle construit est évalué avec les points qui n'ont pas servi à la construction

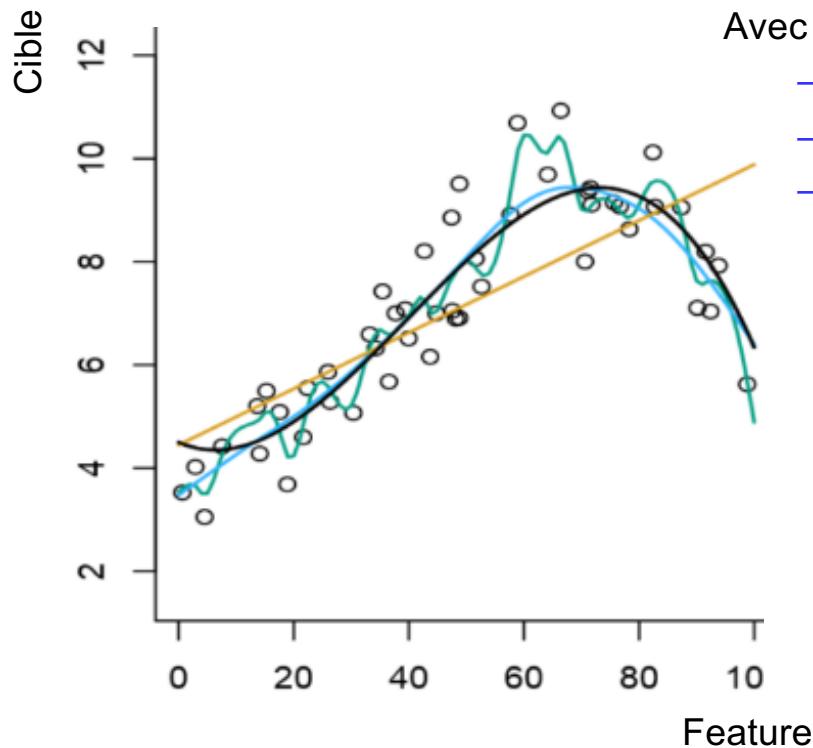


## Une démarche pleine de bon sens 3/3

Ce choix de forme est un a priori, potentiellement guidé par la visualisation des données ou par l'expérience (attention : expérience est une lanterne dans le dos !)

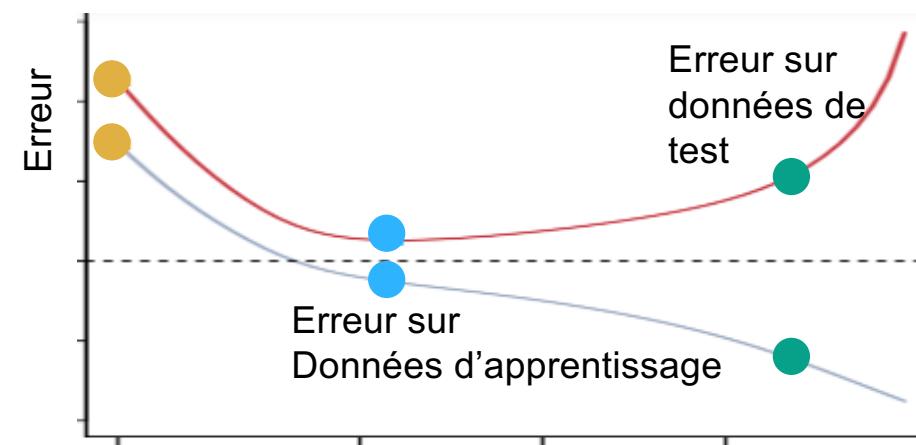


## Zoom : effet de la complexité sur l'erreur de prédiction (1/2)

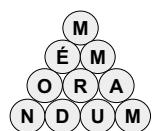


Avec un jeu de données, 3 modèles sont représentés sur ce graphe

- Un modèle complexe : s'approche de près des données observées
- Un modèle plus simple
- Un modèle grossier (ligne droite)



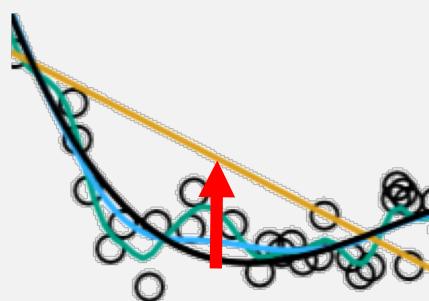
Echelle de complexité du modèle  
(exemple degré de polynôme d'interpolation)



## Zoom : effet de la complexité sur l'erreur de prédiction (2/2)

### Erreurs de biais

- Lié à la « raideur » du modèle
- peu sensible au jeu de données disponibles



Il existe un juste équilibre ...  
et on va le trouver !

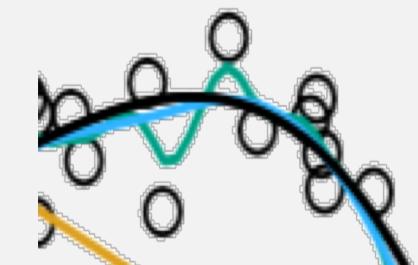


Erreur sur  
données de test

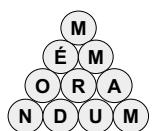
Echelle de complexité du modèle  
(exemple degré de polynôme d'interpolation)

### Erreurs de variance

- Lié à la contingence des données disponibles : juste en moyenne par définition mais toujours versatiles !

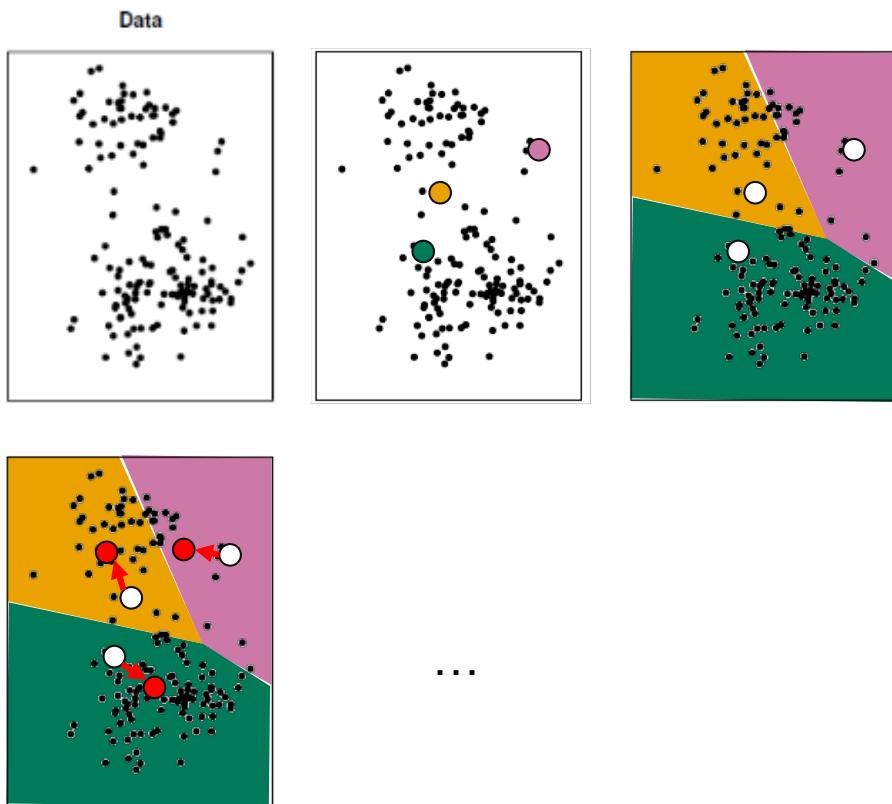


Le bon réglage de la simplification du passé n'est pas fourni par les mathématiques c'est un travail d'artisan !

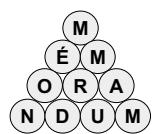
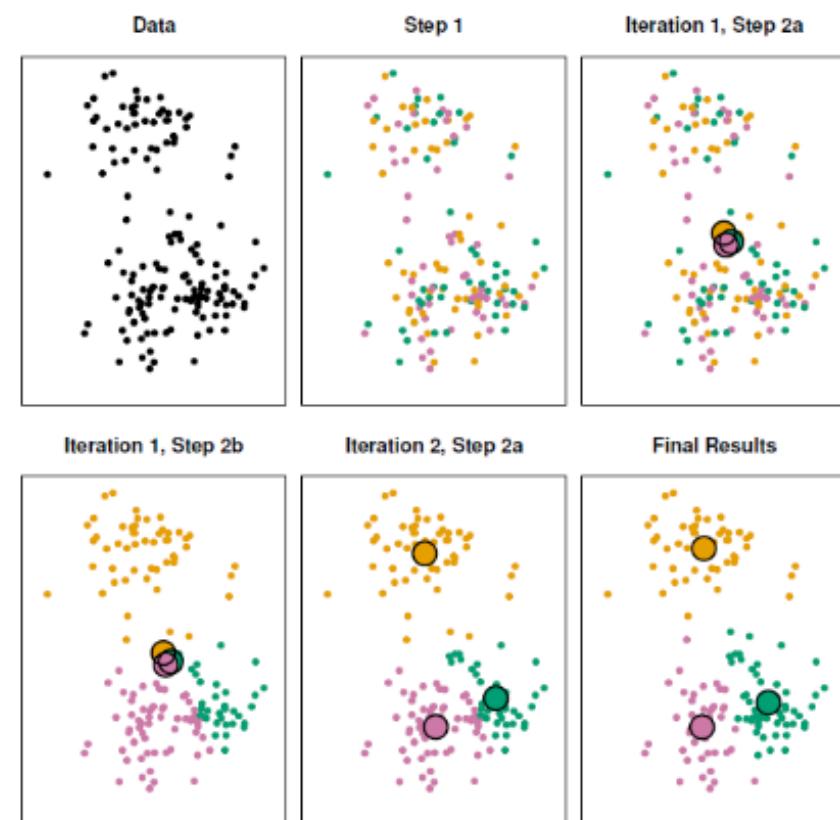


## Et en non supervisé ?

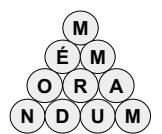
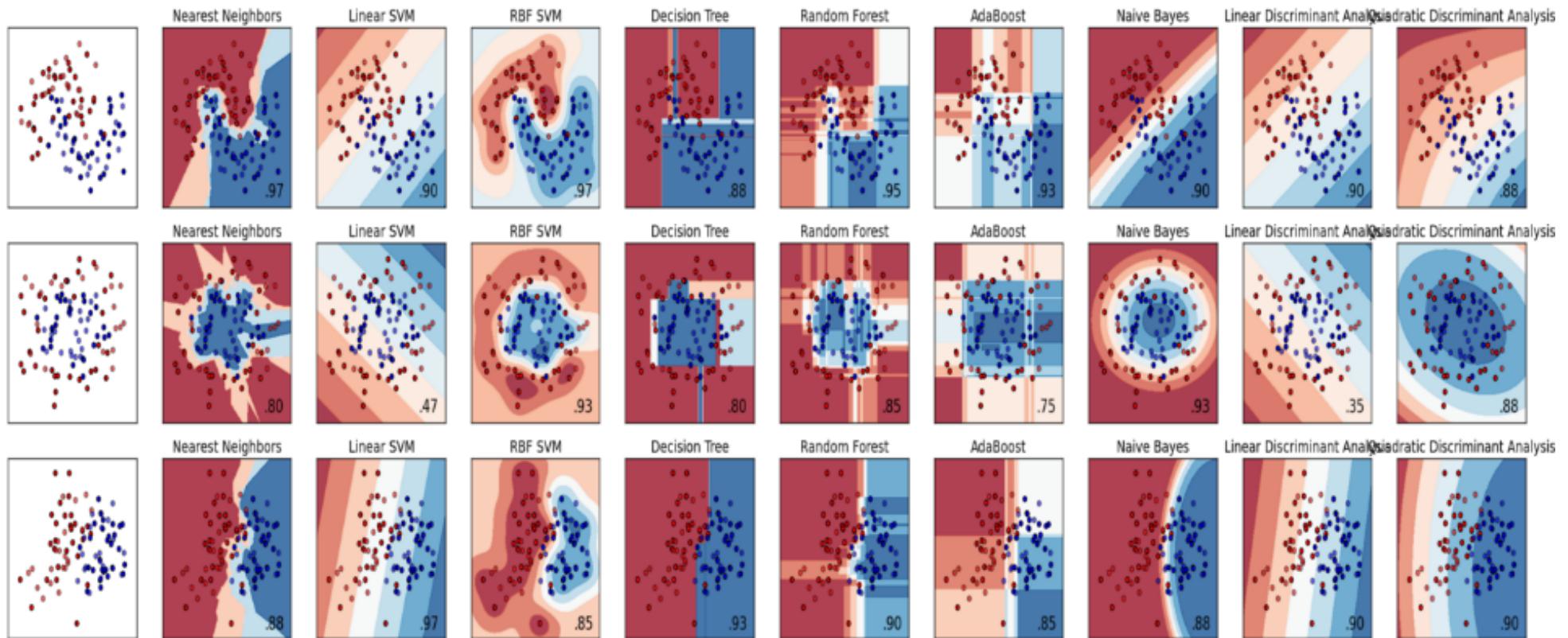
### Choix aléatoire des centroïdes initiaux



### Affectation aléatoire des clusters pour chaque point



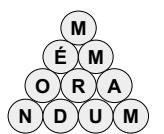
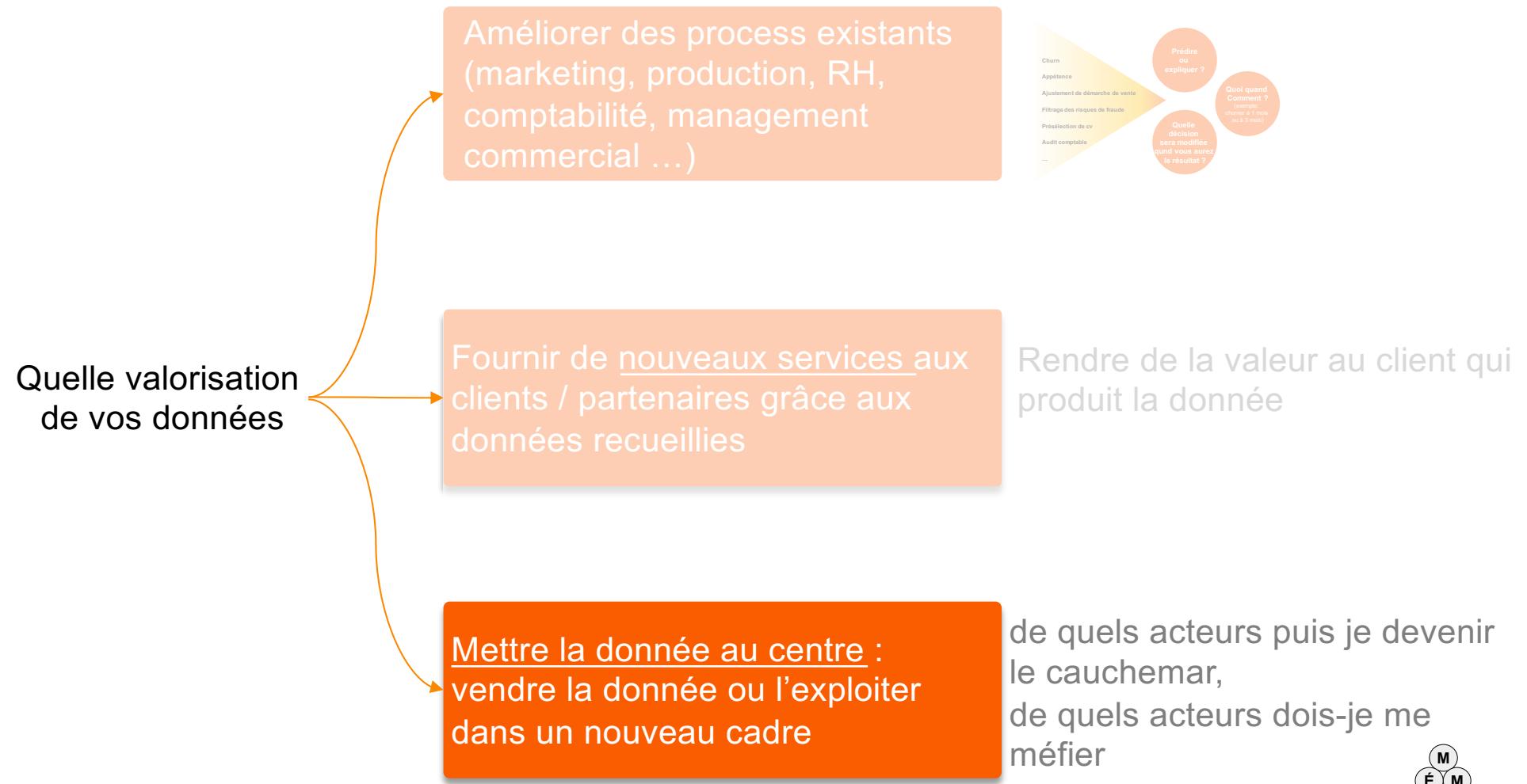
## Une typologie d'algorithme



## L'approche *machine learning* permet de reconstituer un contexte global



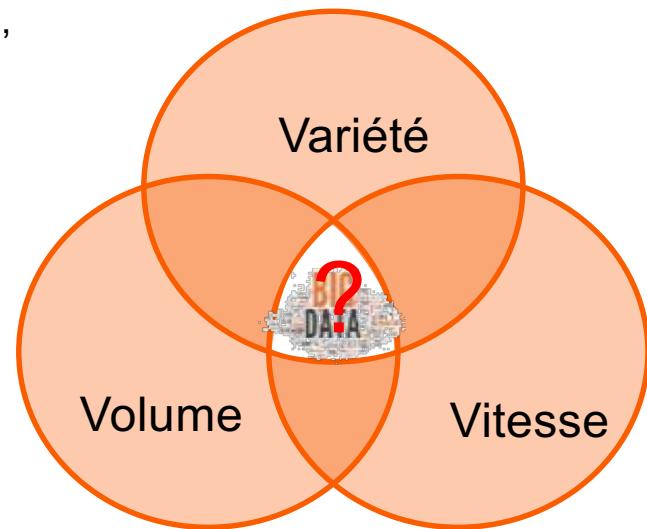
## La donnée peut servir encore dans des contextes très nouveau



## A partir de quel volume est on éligible aux techniques du big data ?

**Une définition fréquente : Au moins 2 des 3 « V »**

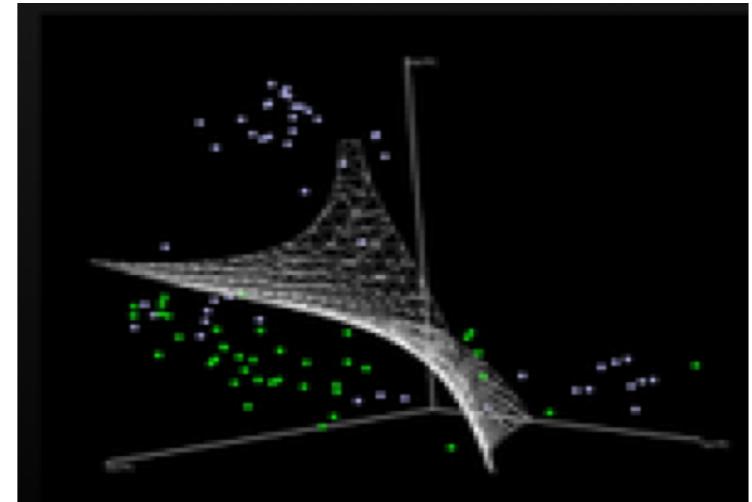
- Volume,
- Vitesse,
- Variété



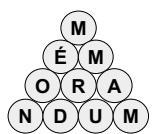
*Les 3V ne sont pas un plancher d'accès mais le symbole d'un plafond constamment repoussé par la technologie*

**Une définition plus juste de la puissance de l'approche → 0 V**

- ci-dessous : 150 observations, 3 caractéristiques → une modélisation fine !



**Votre entreprise a assez de données pour tester l'application des algorithmes prédictifs open source (ceux qui marchent !)**



## *Exemples 1*

### **Entreprise gérant des interventions techniques**

(plusieurs centaines de techniciens sur toute la France)

#### **Contexte initial**

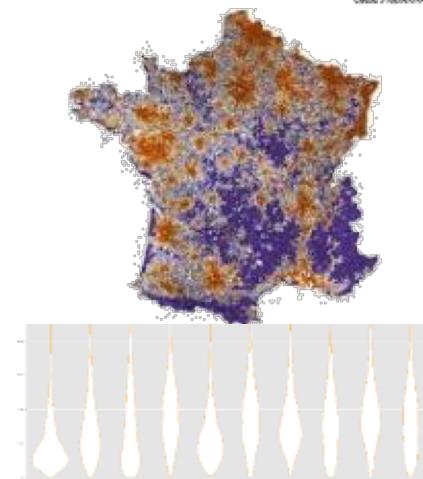
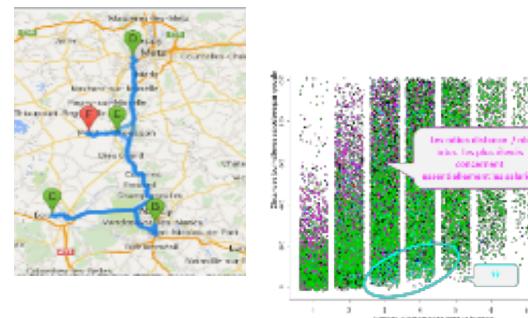
10 ans d'historique  
d'interventions de techniciens,  
jamais exploités

Contexte social tendu sur la  
gestion des planifications de  
tournées

Situation concurrentielle difficile :  
trouver de nouveaux leviers de  
rentabilité

#### **Démarche**

Observation des données et  
entretiens avec le comité de  
direction  
– 6 pistes de travail data



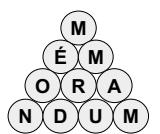
#### **Résultat**

Processus de planification  
refondu (forte évolution des  
principes sous-jacents)

Anticipation des interventions  
difficiles (modèle prédictif)

Aide à la décision sur la  
modification de l'implantation du  
réseau

Identification de nouveaux leviers  
de négociation des contrats  
(modèle prédictif de coûts  
complets)



## Exemple 2

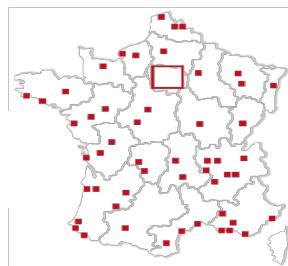
### Réseau de travail temporaire

#### Contexte initial

Un réseau de plusieurs centaines d'agences

Un objectif opérationnel : optimisation de la marge brute par tête

→ Une recherche des leviers d'action cachés



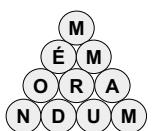
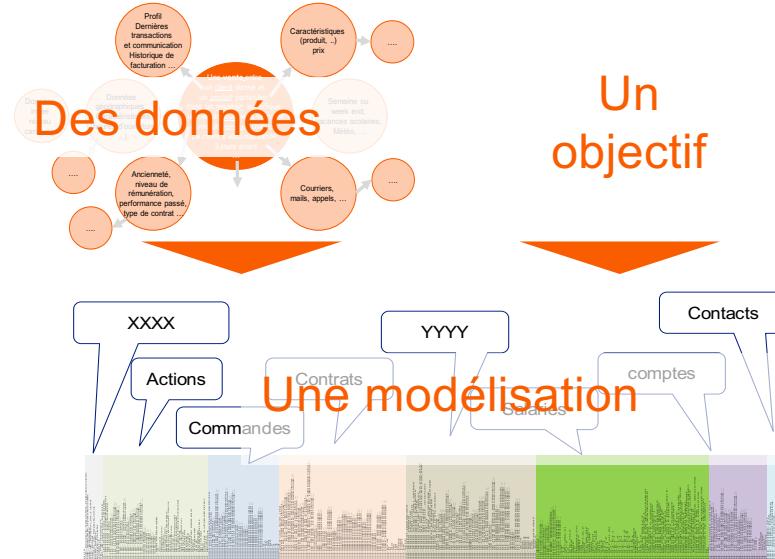
#### Démarche

Constitution d'un large spectre de données opérationnelles (RH/commercial/ marketing / .)  
Modélisation de l'indicateur cible  
Analyse des principaux axes du modèle  
– Les premiers sont évidents  
– Les suivants sont en général une surprise ... le travail commence là



#### Résultat

Exemple : ajuster des heures d'ouverture ou un mode de gestion du flux d'appel en fonction de caractéristiques locales



## Exemples 3

### Filiale assurance vie d'un groupe d'assurance

#### Contexte initial

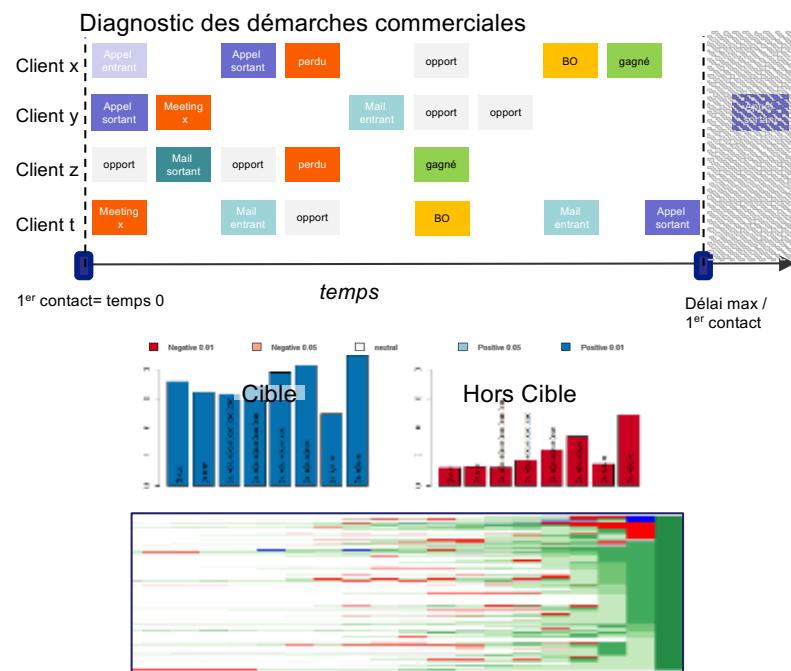
Une jeune équipe de data scientists tout juste formée

Un premier objectif opérationnel sur l'optimisation des forces de vente sur le marché UK



Analyse des séquences commerciales sur les 18 derniers mois (salesforce)

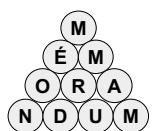
- Visualisation des données
- Modélisation supervisée (probabilités de succès des démarches en cours, faire des recommandations d'action)



#### Résultat

Montée en compétence rapide de l'équipe (acquisition de données, construction d'interface graphiques, présentation de résultats actionnables)

Prise de conscience par les managers commerciaux de la réalité des démarches commerciales



## Exemples 4

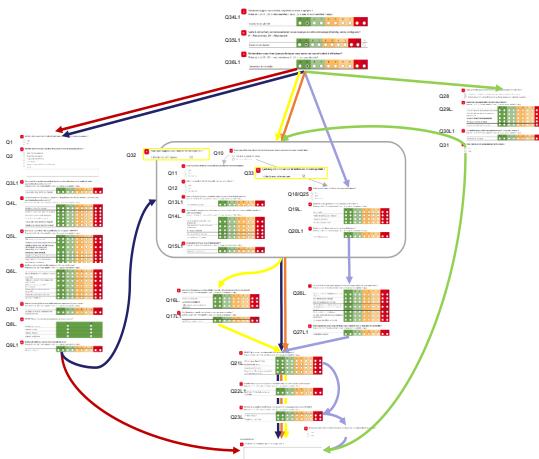
### Opérateur d'enquête de satisfaction client

#### Contexte initial

Des questionnaires de satisfaction administrés au fil de l'eau

Des centaines de séries temporelles à interpréter !

- Par magasin
- Par question
- Par type de client



#### Démarche

Identifier les profils opérationnels, leurs objectifs spécifique (retour du client, rentabilité, ..) et leurs leviers d'action

Modéliser les objectifs de chacun : identifier les indicateurs de satisfaction en lien direct avec leurs objectifs

Transformer les données de satisfaction en recommandations opérationnelles pour chaque profil

**Evaluer en amont les objectifs de chaque profil interne  
Exemple centre de vacances**



##### Profil

Directeur régional

Directeur de site

Chef d'étage NPS

##### Objectif

rentabilité

Retour client

NPS

##### Exemple leviers

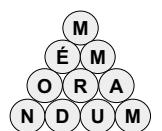
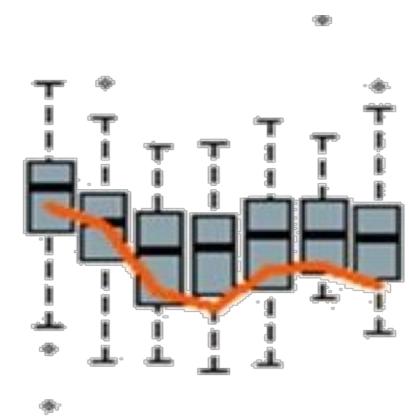
Organisation

Parcours d'accueil

Propreté, ...

#### Résultat

Un outil de management opérationnel



## Exemples 5 Centres de contacts → une mine !

Mieux évaluer la performance des conseillers

Événement précédent du client



Contacts du conseiller X

Événement suivant du client

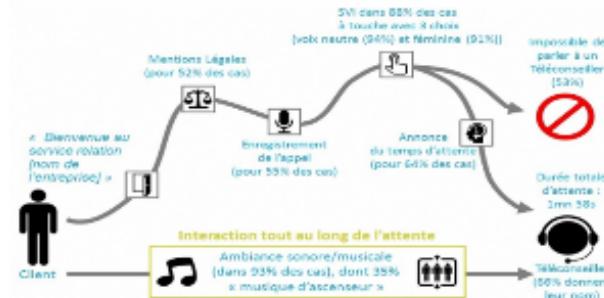


AdAPTER les démarches de vente au client en ligne

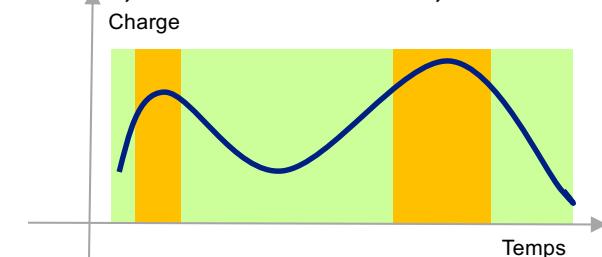
*Le client est en fin de droit et ...*  
*Le client est débile et ne comprends rien à ce que je lui explique*  
*Le client n'a pas de ressources*  
*Le client XX est en prison pour X années*  
*La discussion avec le client n'a pas permis d'aboutir*  
*La discussion est difficile ....*

Identifier et reformuler des commentaires désobligeants sur les clients

Améliorer la parcours de prise en charge des clients



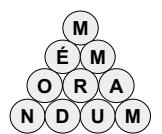
Prévision de flux intégrant le trafic web, la météo, les match de foot, ..



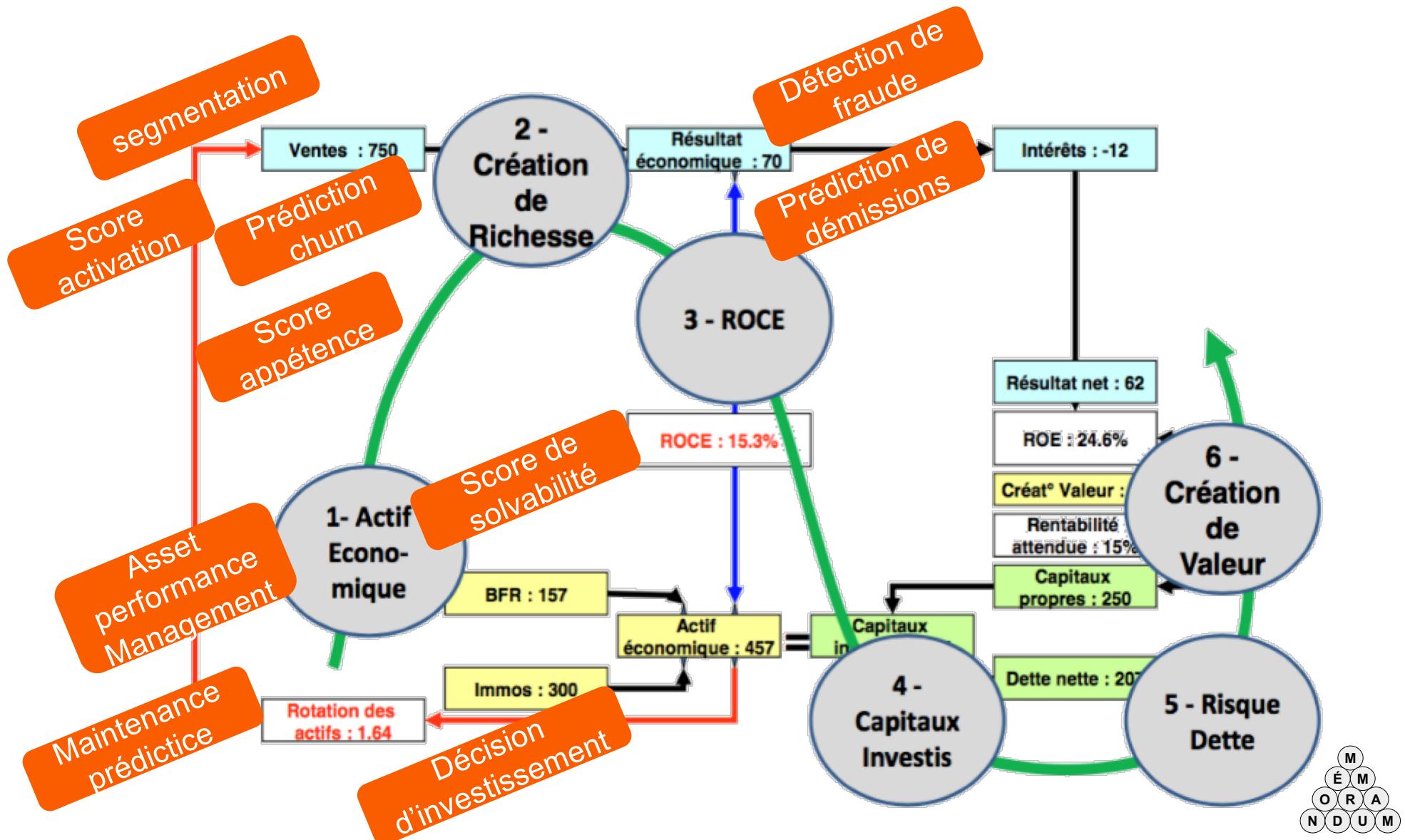
Identifier les motifs corrélés aux pics de charge : les anticiper



Identifier les thématiques abordées par les clients

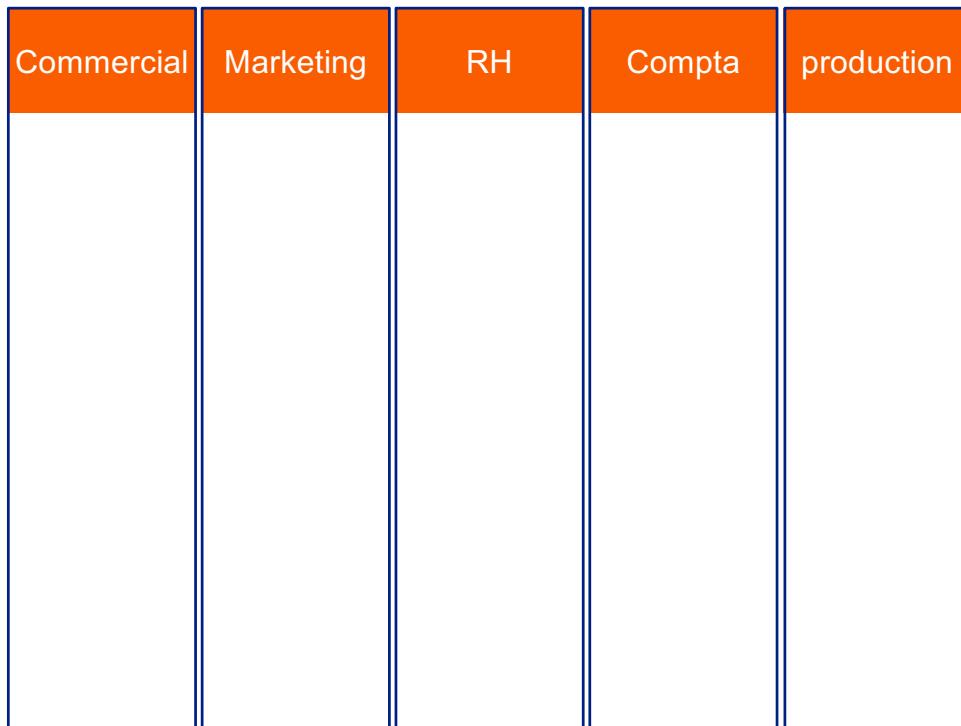


## Des cas d'usage qui suivent tout le modèle économique et financier



## Managez vos données comme vous managez vos équipes

**Vos données sont rangées comme vos organisations**



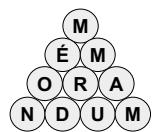
**Vos données doivent être animées comme votre management : transverse**



**Pas d'un projet informatique mais de management. Objectif managers :**

- Être garant de la disponibilité de ses données
- réutiliser les données des autres directions

**Vos données ne travailleront ensemble que sous l'impulsion de votre  
Comité de Direction**



## Plus globalement La bonne recette

1

### « La valeur n'attend pas le nombre des octets »

Une valorisation efficace des données dépend moins du nombre d'observations (quelques milliers de clients ou de dossiers représentent déjà une bonne base) que de la richesse de ses observations (pour un client avoir ses caractéristiques, ses transactions, ses courriers, ...)

2

### « Exploiter la donnée est un projet métier, pas un projet informatique »

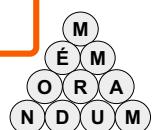
L'initialisation d'une démarche de valorisation des données ne nécessite aucune infrastructure :

- Des ordinateurs de bureau (jusque quelques giga de données)
- Des logiciels d'analyse open source (R / Python)
- Des algorithmes d'analyse libres de droit, prédéveloppés et prêt à l'emploi

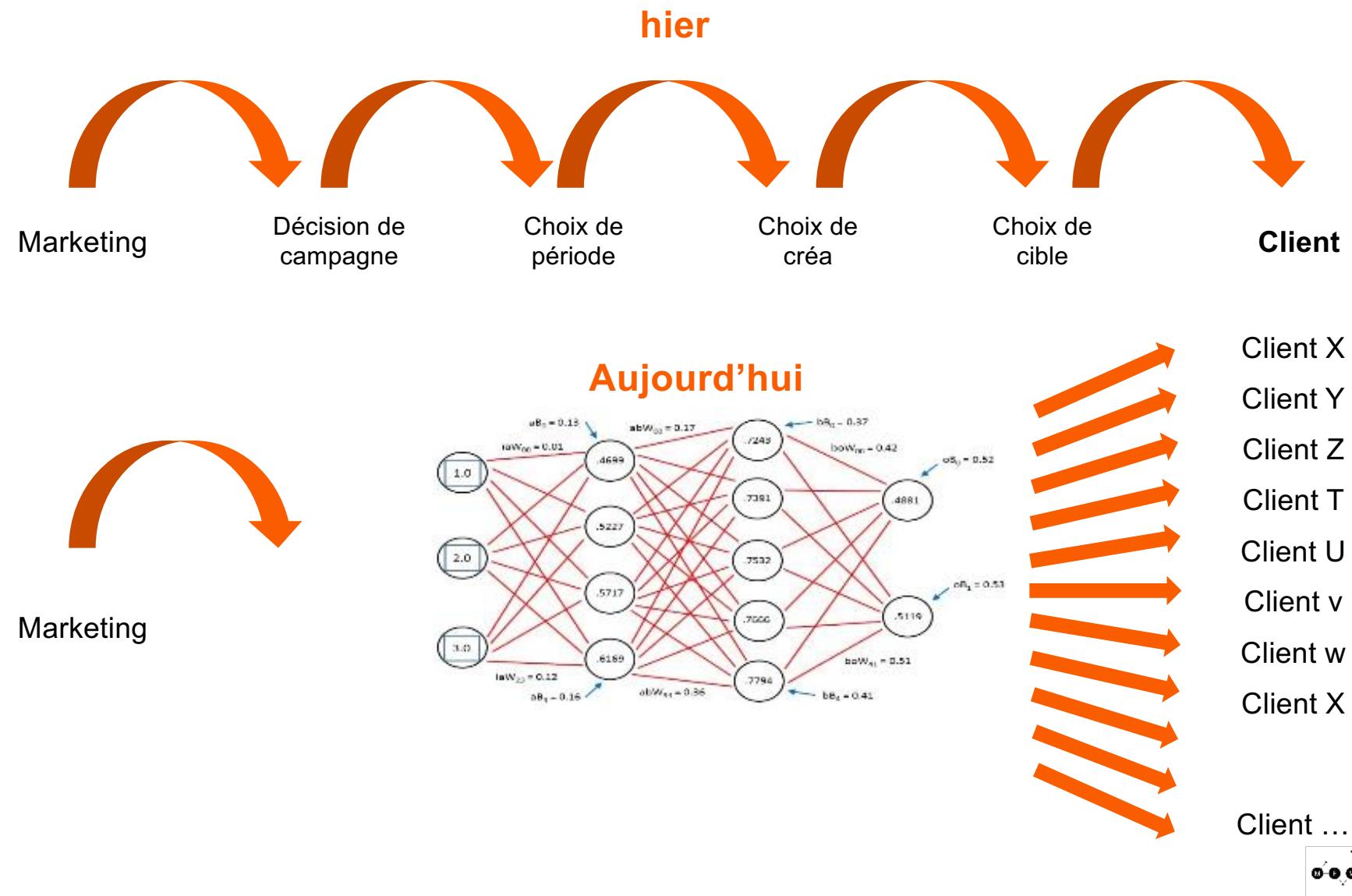
3

### « Fail fast & learn »

Les démarches de valorisation des données se font par itérations, il est donc important de savoir avant de se lancer que certaines analyses potentiellement non concluantes (mais auront fait progresser)



## Un circuit de décision transformé



## L'exploitation des données internes de l'entreprise cumule 3 avantages

1) Travail sur des données naturellement riches et accessibles

2) Pertinentes pour les algorithmes d'analyse

3) Sans besoin d'infrastructure

