

# Intelligence artificielle et Machine Learning

Concepts, Méthodes, Normalisation

Aurélien Garivier

École Normale Supérieure de Lyon, UMPA & LIP

Jeudi 2 décembre 2021



# Ordinateur et problèmes

**Ordinateur** = machine capable d'un petit nombre d'opérations élémentaires, qu'on peut combiner à l'envi

**Spécification d'un problème** = description de la sortie souhaitée pour une entrée donnée



Exemples:

[3,2,5,1,4]

→

[1,2,3,4,5]

le petit chat

→

the little cat



→

5



→



# Deux approches de résolution

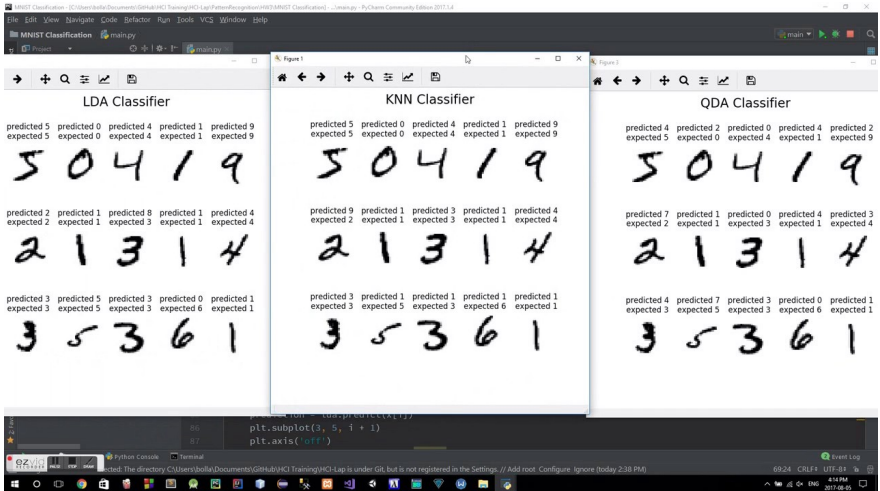
**Approche classique : réduction** = décrire le *programme* = la suite d'opérations élémentaires permettant de construire la sortie à partir de l'entrée  
= programmation (codage)

**Intelligence artificielle** = programmer l'ordinateur pour qu'il *construise lui-même* le programme qui résout la tâche visée  
= méta-programmation

→ pour chaque problème, les deux approches sont possibles

→ mais elles sont plus ou moins efficaces

# Exemple tarte à la crème: MNIST



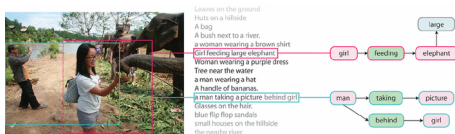
# Des succès spectaculaires

- Reconnaissance d'image
- Traitement du langage naturel
- ... et leur combinaison

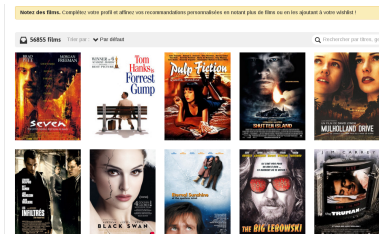
- Résolution de jeux (stratégie)
- Véhicules autonomes



- Systèmes de recommandation (presse, publicités, films, etc.)



<https://link.springer.com/article/10.1007>



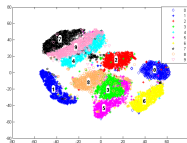
# Science des données

= résolution de problèmes dont les entrées sont les données (singulier politique)

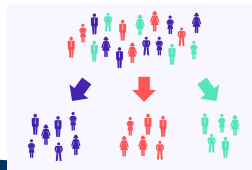
→ stockage, gestion, accès, etc.



→ visualisation, représentation




→ simplification, typologie



# Intelligences Artificielles

→ **Raisonnement sur les données**

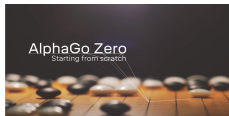
Ontologies  
Web sémantique

Ex:  + Wikipedia → Léonard de Vinci

→ **Stratégies**

Jeux  
Agents autonomes

Ex:



→ **Apprentissage automatique**

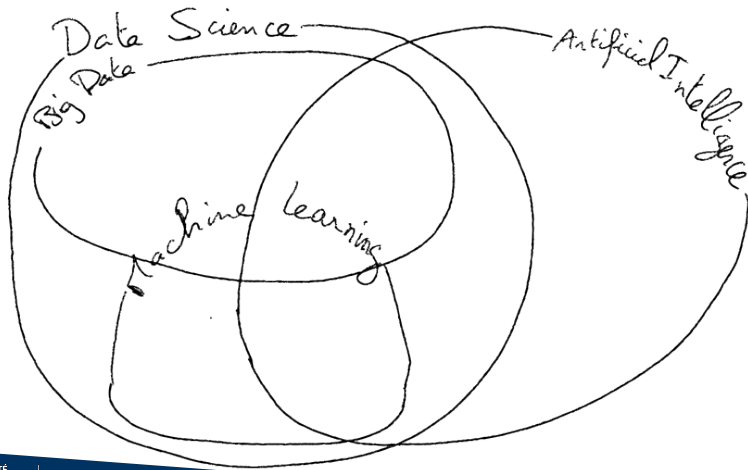
Classification supervisée  
Régression

Ex:



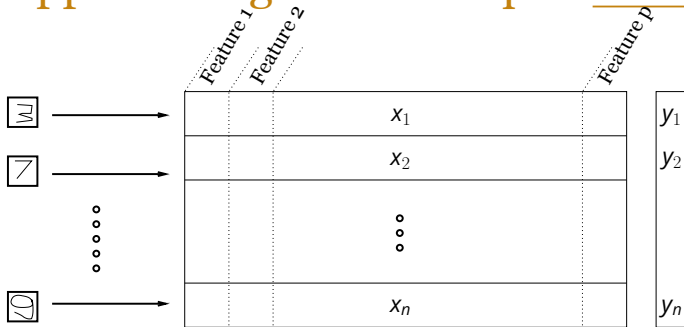
→ chat

# Science des Données, Intelligence Artificielle, Apprentissage Automatique





# Apprentissage automatique: normalisé



$$X \in \mathcal{M}_{n,p}(\mathbb{R})$$

$$Y \in \mathcal{Y}^n$$

Données : matrice  $n \times p$

- $n$  exemples indépendants
- $p$  features = caractéristiques mesurées

Classifier  $\mathcal{A}_n$

$$\begin{array}{c} \downarrow \\ h_n : \mathcal{X} \rightarrow \mathcal{Y} \\ \boxed{6} \mapsto 6 \end{array}$$

# Normalisation d'un problème pour l'apprentissage automatique

Chaque exemple est un vecteur

ex: TAL: phrase = "sac de mots"

Les exemples sont indépendants

ex: séries temporelles

Ils couvrent tous les cas de figure

extrapolation très hasardeuse

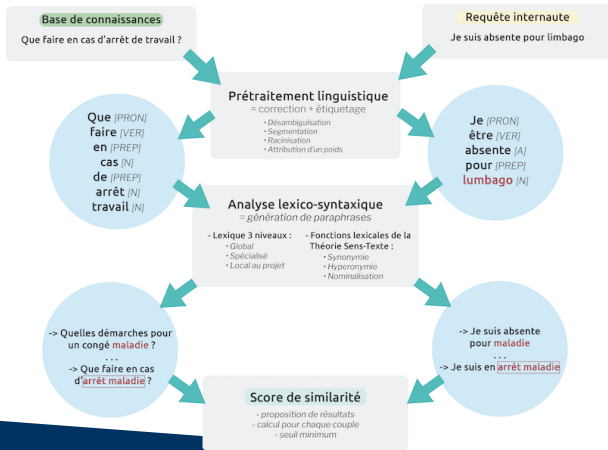
# Exemples

- Catégorisation des pages: sommaire / page blanche / etc.  
ML ou modèle stochastique avec dépendance ?
- Amélioration des images (correction couleurs, super-résolution etc.)  
traitement d'image? ou apprendre les traitements à base d'exemples ?
- Segmentation des pages / étiquetage des zones  
règles explicites ou à apprendre ?
- Etiquetage des illustrations  
ontologies ?
- Identification des identités nommées pour une bibliographie  
construction en ligne de catégorisations ?

# Exemple : Création automatique de FAQ

inbenta

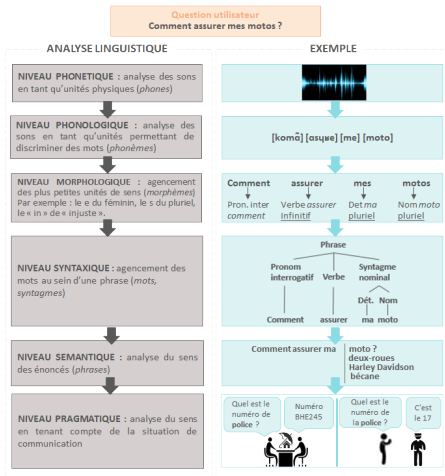
## Fonctionnement de notre moteur



UNIVERSITÉ  
DE LYON



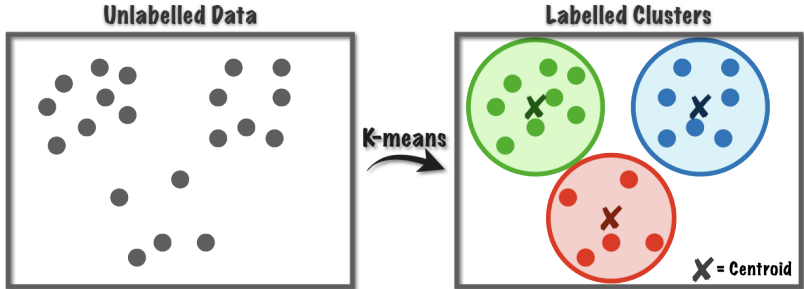
ENS DE LYON



Src: Inbenta.com

# Projet spécifique

- Compétences linguistiques fortes
- Besoin d'algorithmes personnalisés pour le contexte précis
- Création automatique d'une FAQ
- Enormément de questions, à traiter automatiquement
- Besoin d'énormément de robustesse (beaucoup de "bruit")



# La solution

- Les connaissances linguistiques donnent une excellente notion d'affinité entre mots
- On l'étend intelligemment à une bonne notion de distances entre questions
- Puis on met en œuvre une stratégie d'apprentissage non-supervisé basée sur les affinités
- k-means est identifiée comme la plus prometteuse, mais
  - trop chère en coût de calcul
  - trop sensible au bruit
- Une variante est développée qui est paramétrable pour les besoins spécifiques