

**UNIVERSITÉ
TOULOUSE
CAPITOLE**



Intelligence artificielle Machine Learning

**Benoit Gaudou
Moncef Garouani**

Plan du cours

- Introduction au Machine Learning
- Introduction à l'apprentissage supervisé
 - Notions et définitions
 - 5 algorithmes:
 - K plus proches voisins (kNN)
 - Régression linéaire et logistique
 - Support Vector Machine (SVM)
 - Arbre de décision (Decision Tree)
 - Forêts aléatoires (Random Forrest)
 - Evaluation des modèles
 - Optimisation paramétrique
 - Point méthodologique

Organisation et Evaluation du cours IA - Machine Learning

- **Heures**

- 6h de CM (9/02 + 16/02)
- 6h de TP (18-19/2 + 04-05/03)
- Evaluations : 1h examen papier sur table (19/03) + 2h de TP noté (12/03)

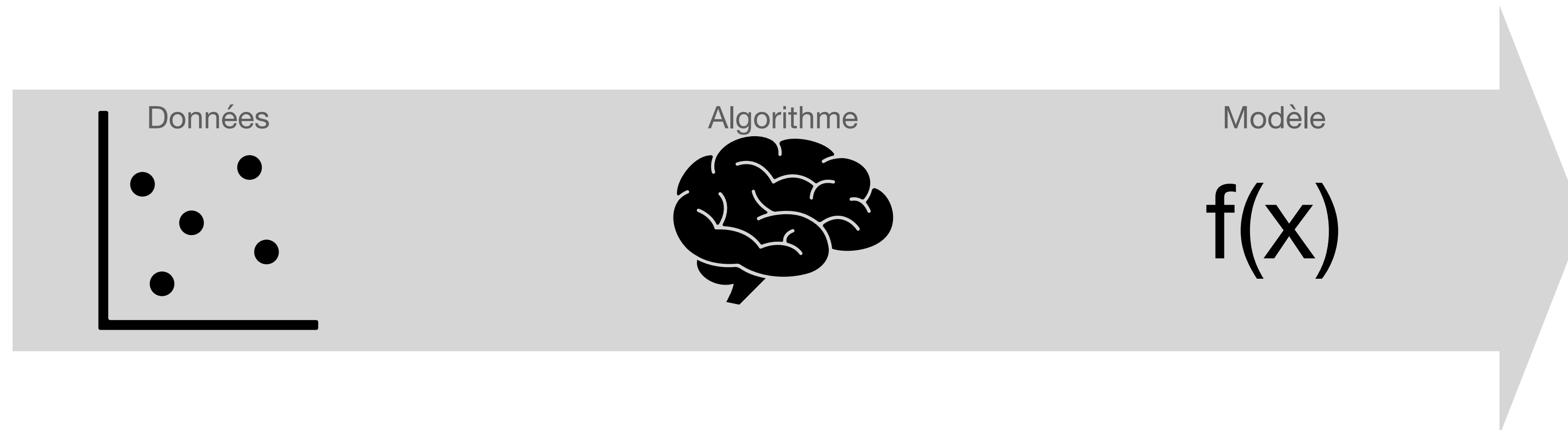
- **Evaluations**

- Examen papier (aucun document autorisé) - 50% (de la note d'IA-ML)
- TP noté : sur machine de l'université, Google Colab
- Possibles controles/QCM/quizz surprises

- **La note du cours d'Intelligence Artificielle = moyenne de la note IA-ML et IA-Algorithmme-Résolution-Problème**

Qu'est-ce que le machine learning ?

- Programme basé sur les données
 - Capacité des ordinateurs à **apprendre à réaliser une tâche** (classification, détection, traduction, etc.) **sans l'avoir explicitement programmé**
 - Etude des algorithmes qui **améliorent leur performance** sur **certaines tâches** en se basant sur leur **expérience**



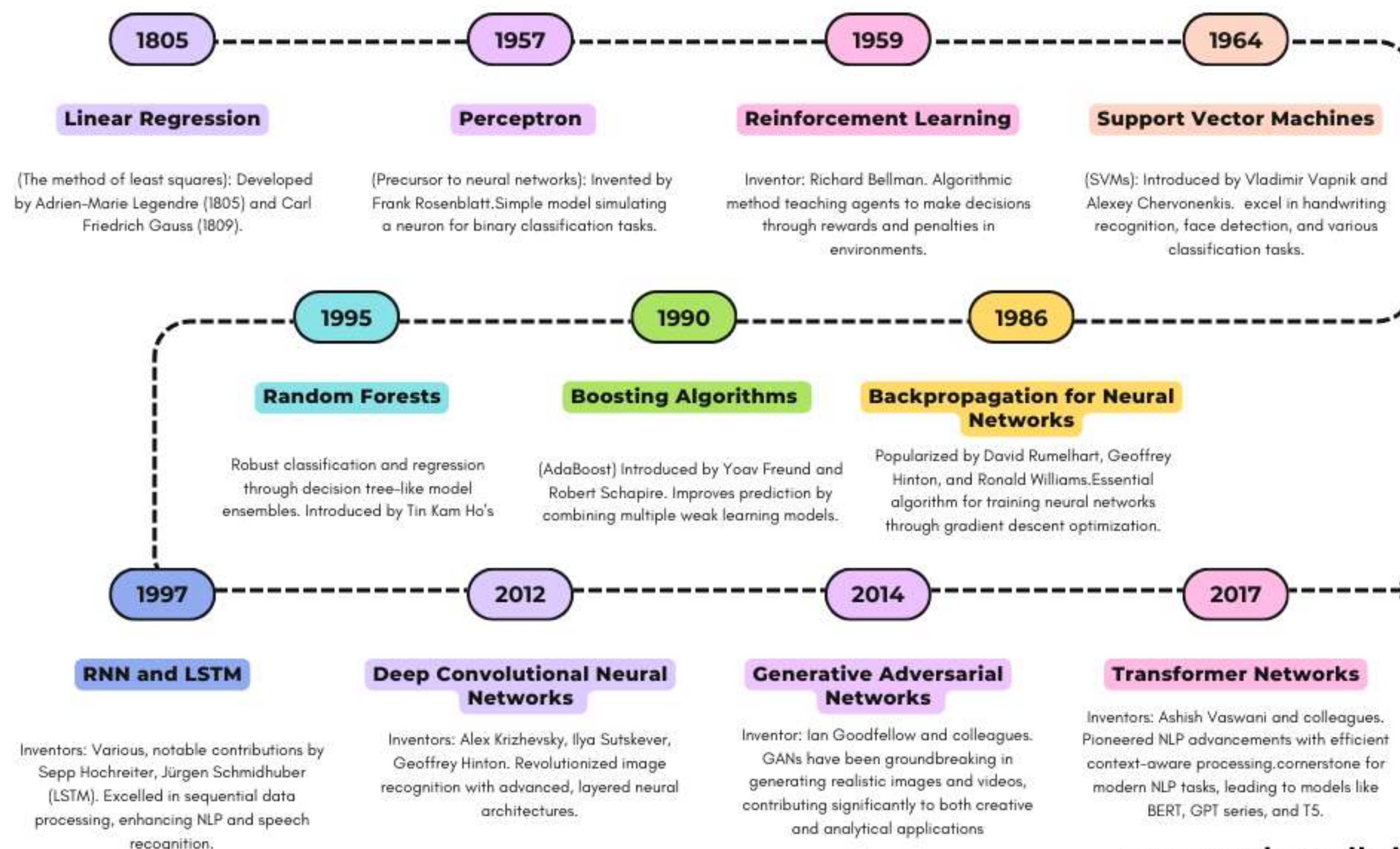
Qu'est-ce que le machine learning ?

Historique

- Le terme **apprentissage machine** a été introduit **1959** grâce par **Arthur Samuel**.

Programme capable d'apprendre à jouer aux dames.

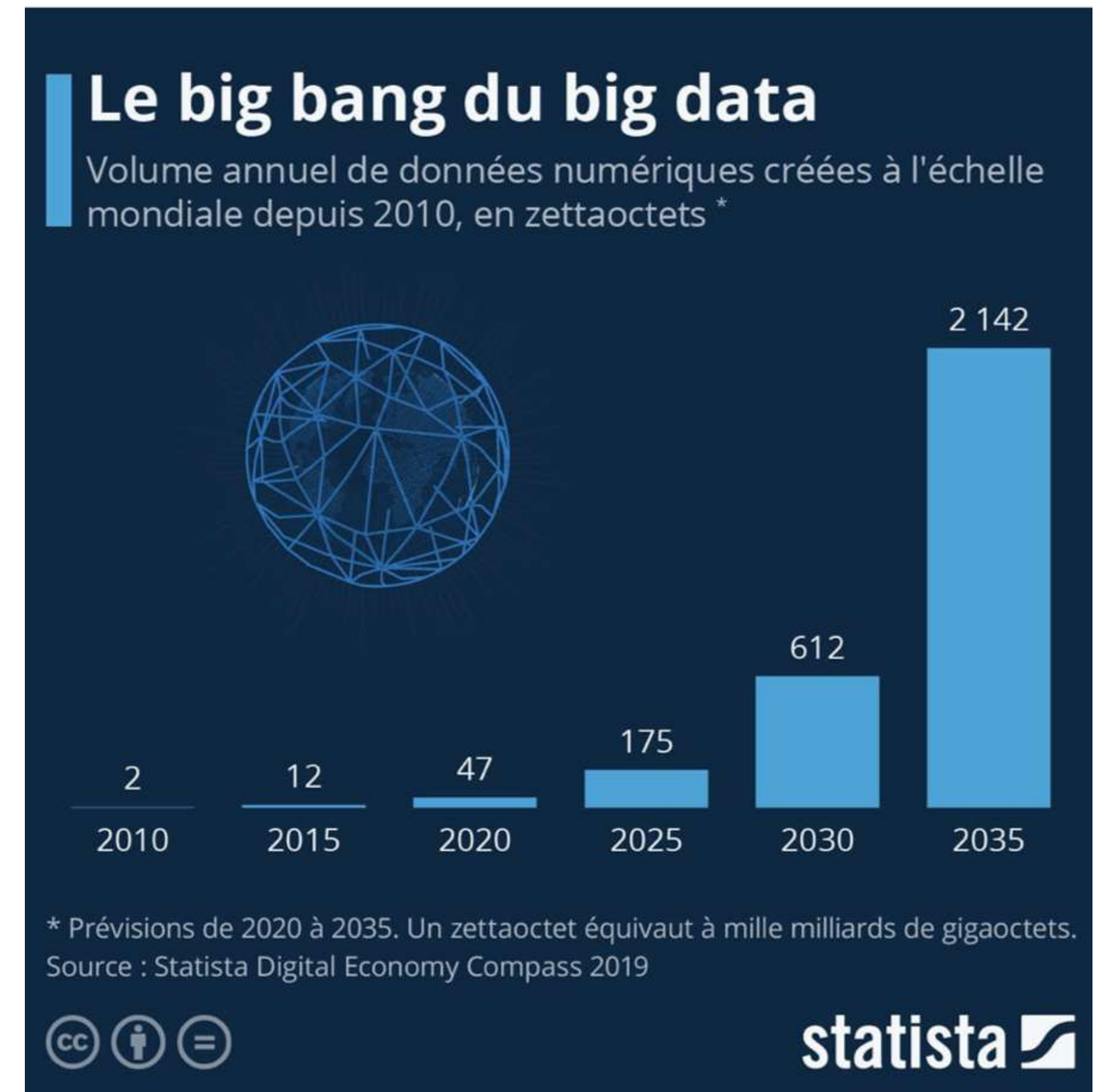
1755-57 : Ruđer Josip Bošković



Pourquoi le Machine Learning ?

Pourquoi maintenant ?

- Disponibilité des données
 - Aire du Big Data : augmentation continue des données générées
 - X : 500 M tweets / jour (= 50 à 140To/jour)
 - Facebook : ~4 Po / jour
 - Youtube : 720 000h de vidéos uploadées par jour (~4 Po / jour)
 - 251 millions d'emails par minute (361 milliard / jour)
 - Données de capteurs : Internet of things... (~73 zettaoctet en 2025)



Pourquoi le Machine Learning ?

Pourquoi maintenant ?

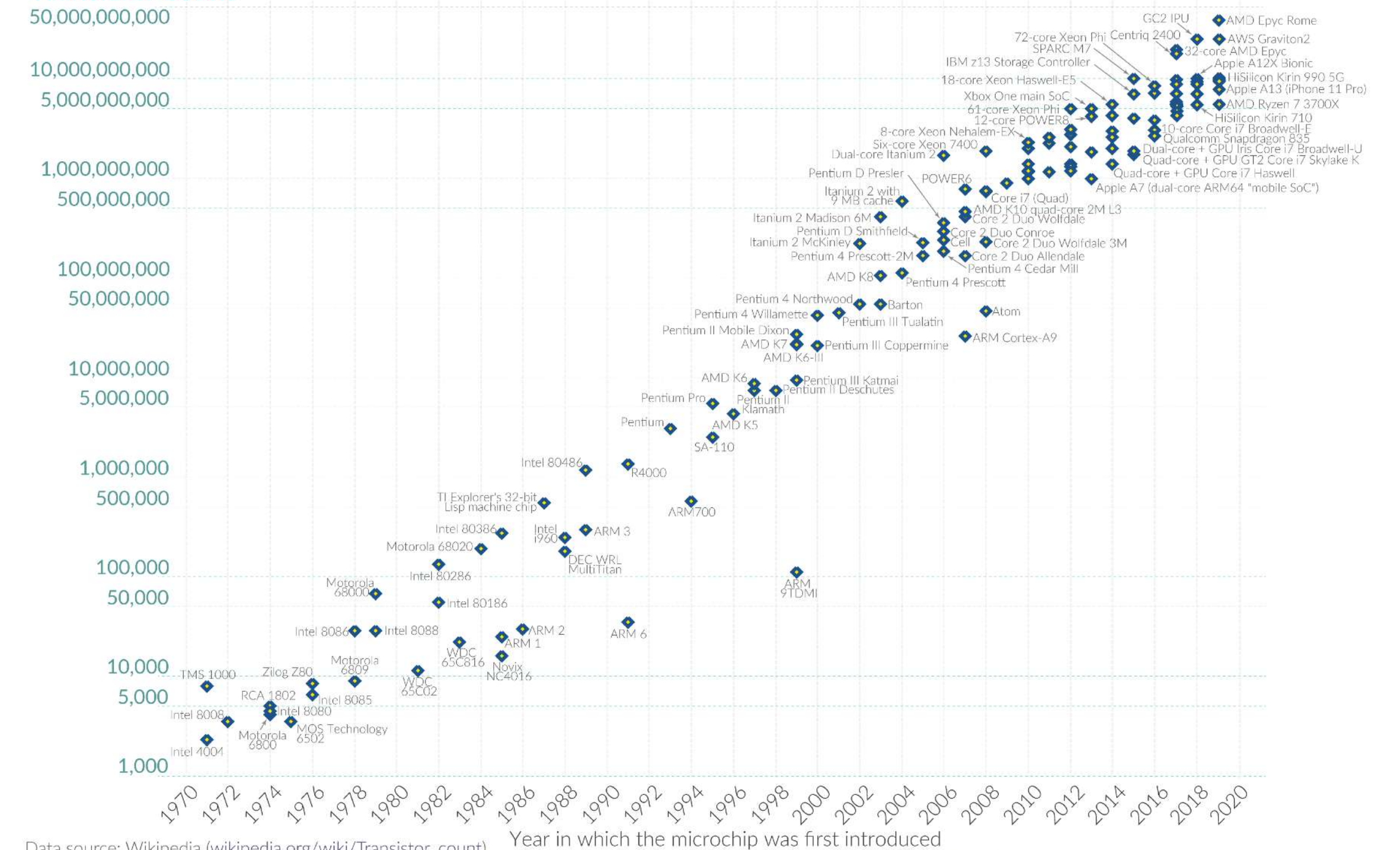
- Augmentation de la puissance de calcul
- Loi de Moore : doublement de la puissance tous les 2 ans
- Calcul parallèle et matriciel

Moore's Law: The number of transistors on microchips doubles every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Our World
in Data

Transistor count



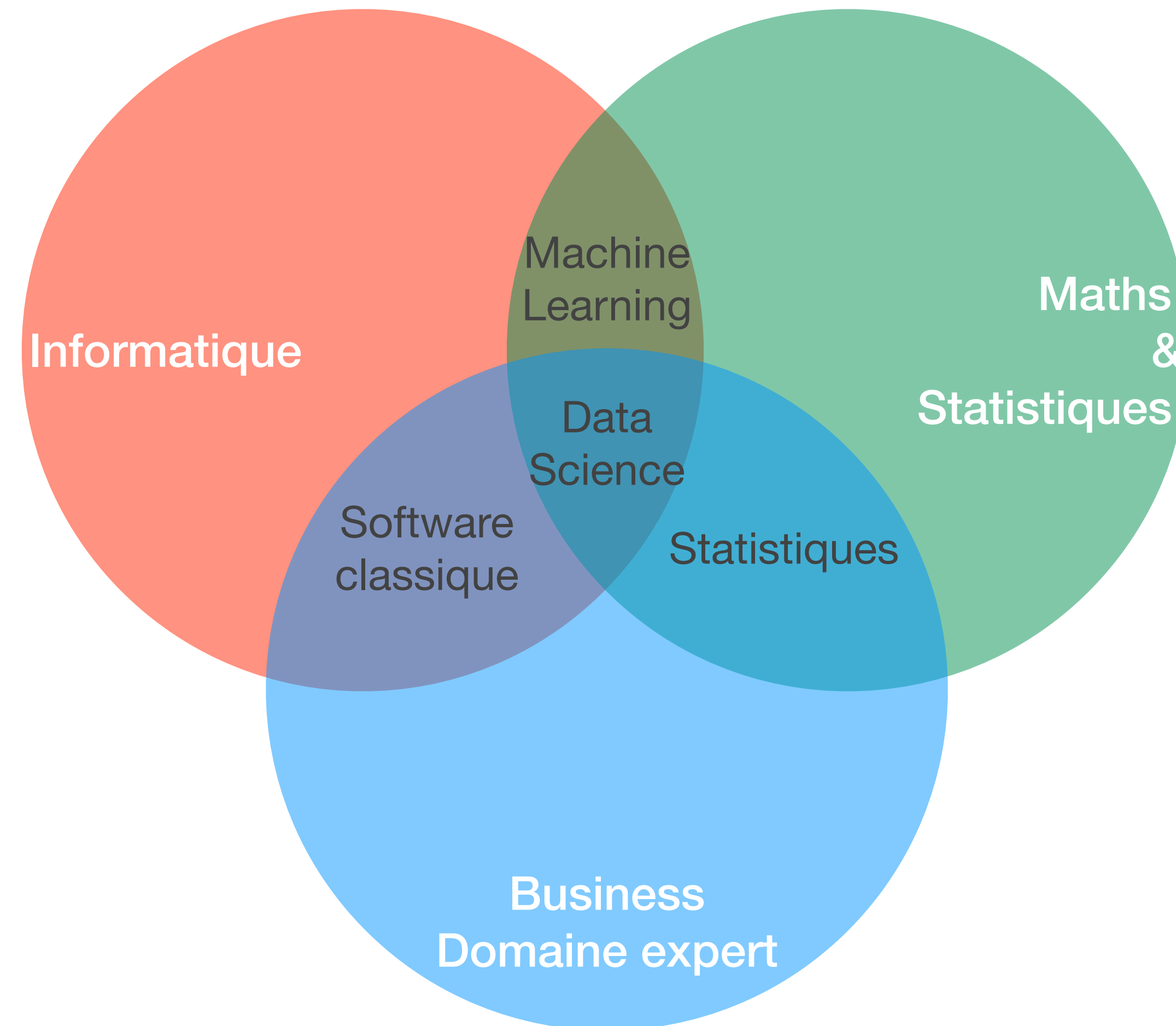
Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)

OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.

Positionnement du Machine Learning

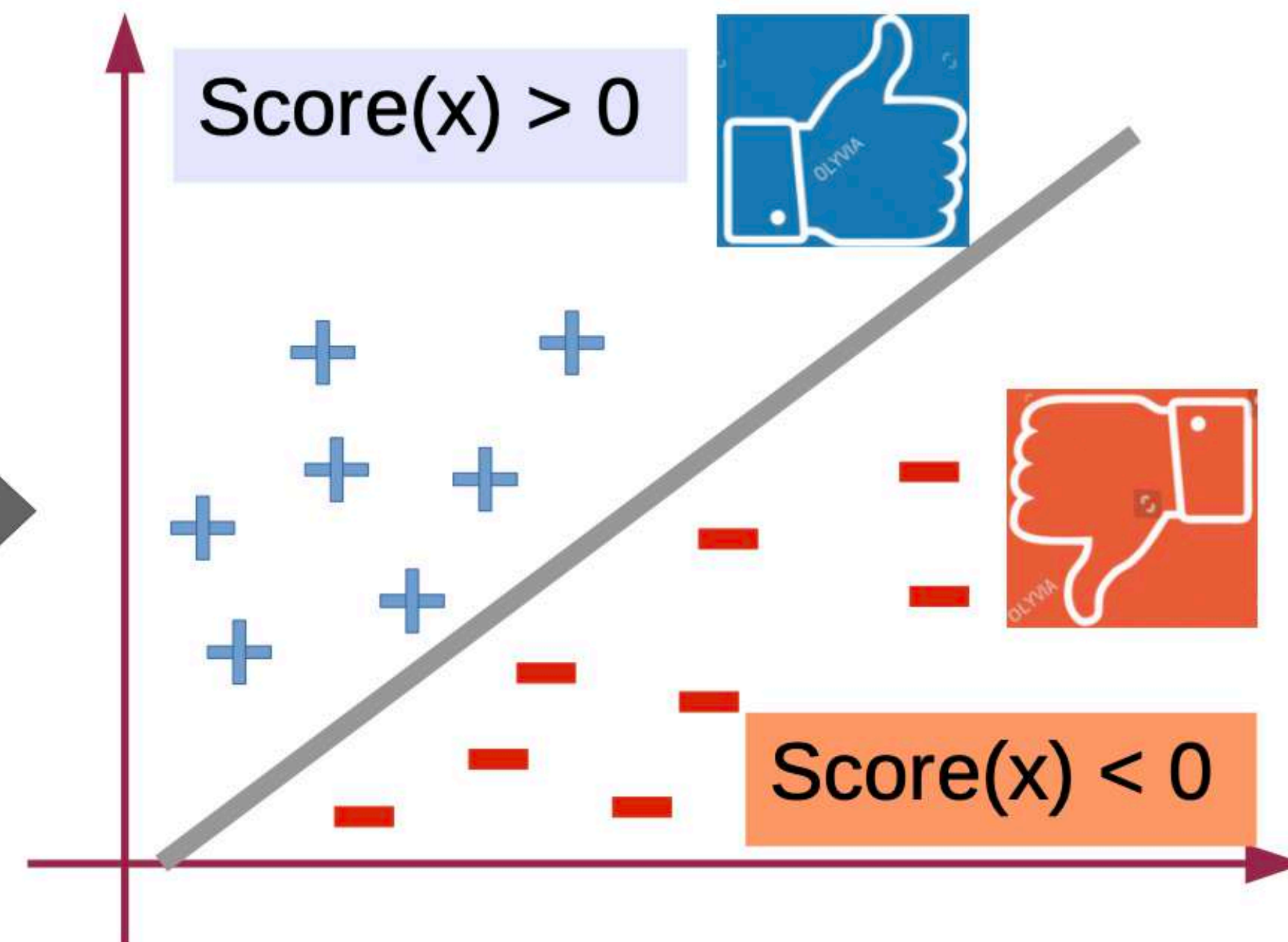
Et Data science



Application du machine learning

Analyse des sentiments

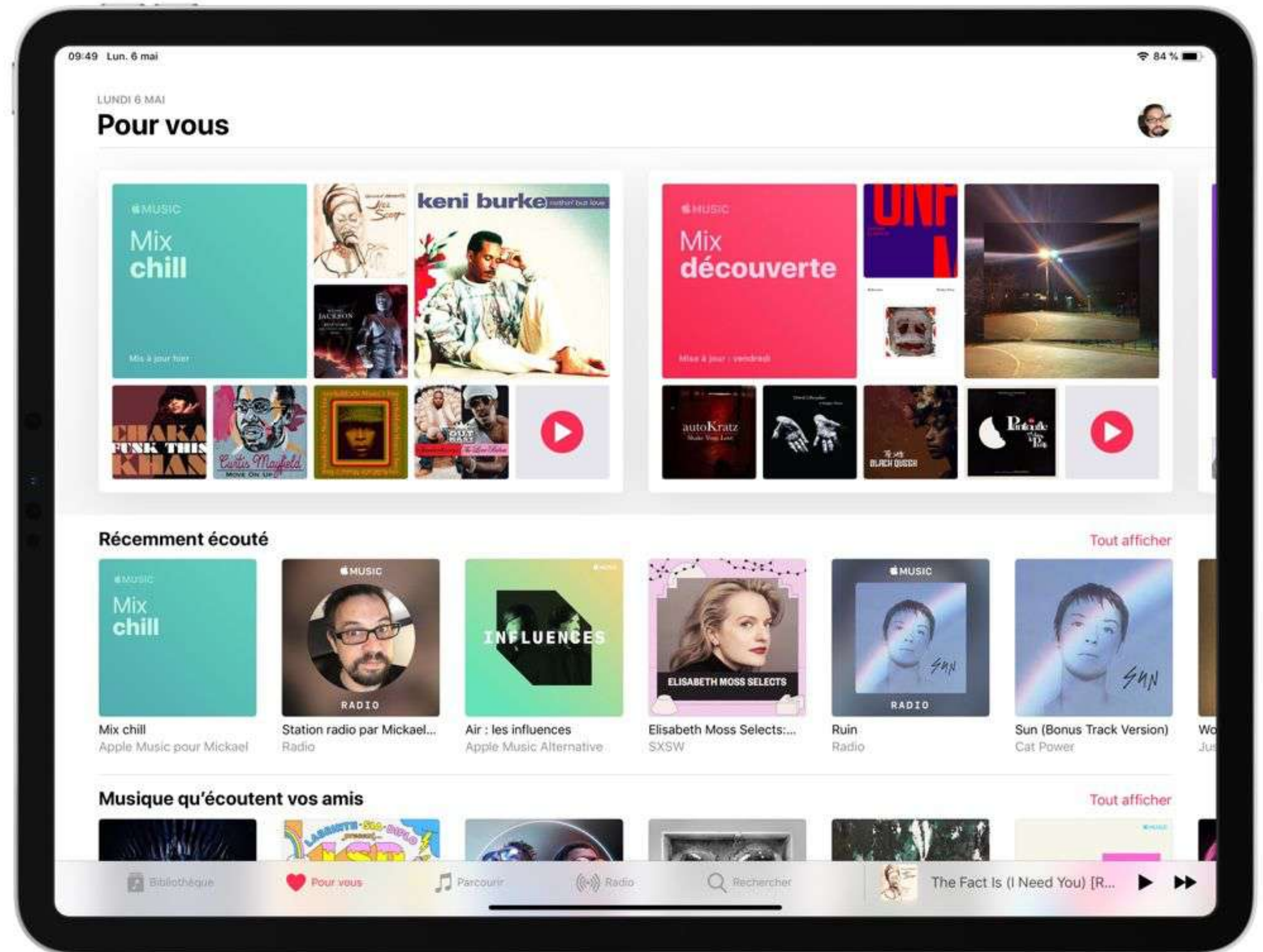
- Classifier les commentaires des utilisateurs



Application du machine learning

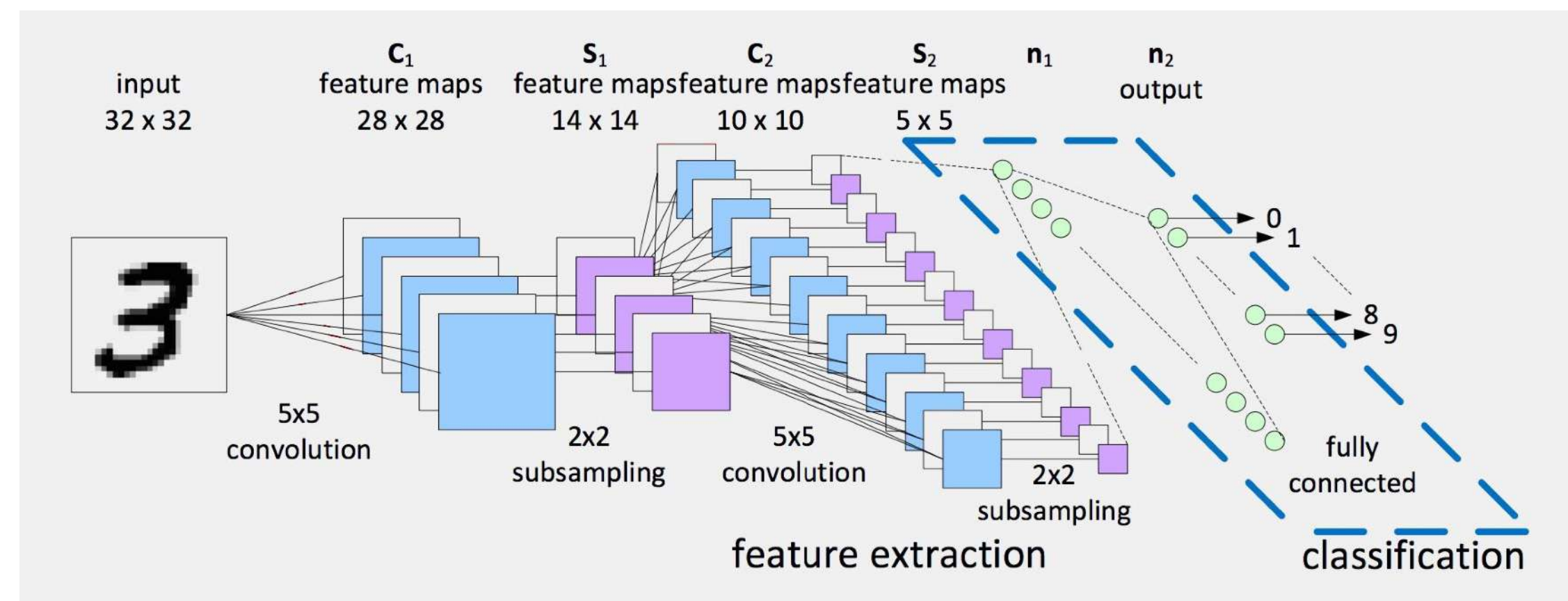
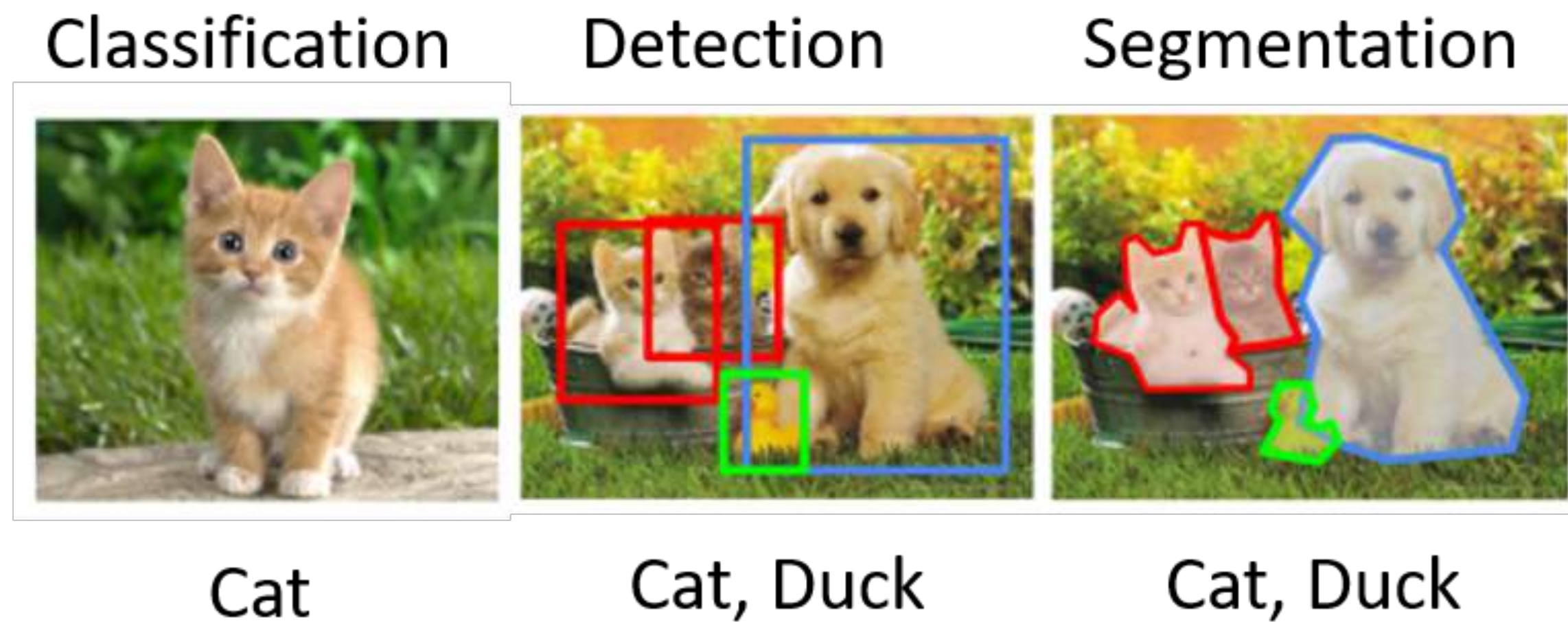
Recommandation de produit

- Proposer de nouveaux produits en fonction des caractéristiques des personnes
- Données réseaux sociaux,
- Historiques de navigation,
- Etc.

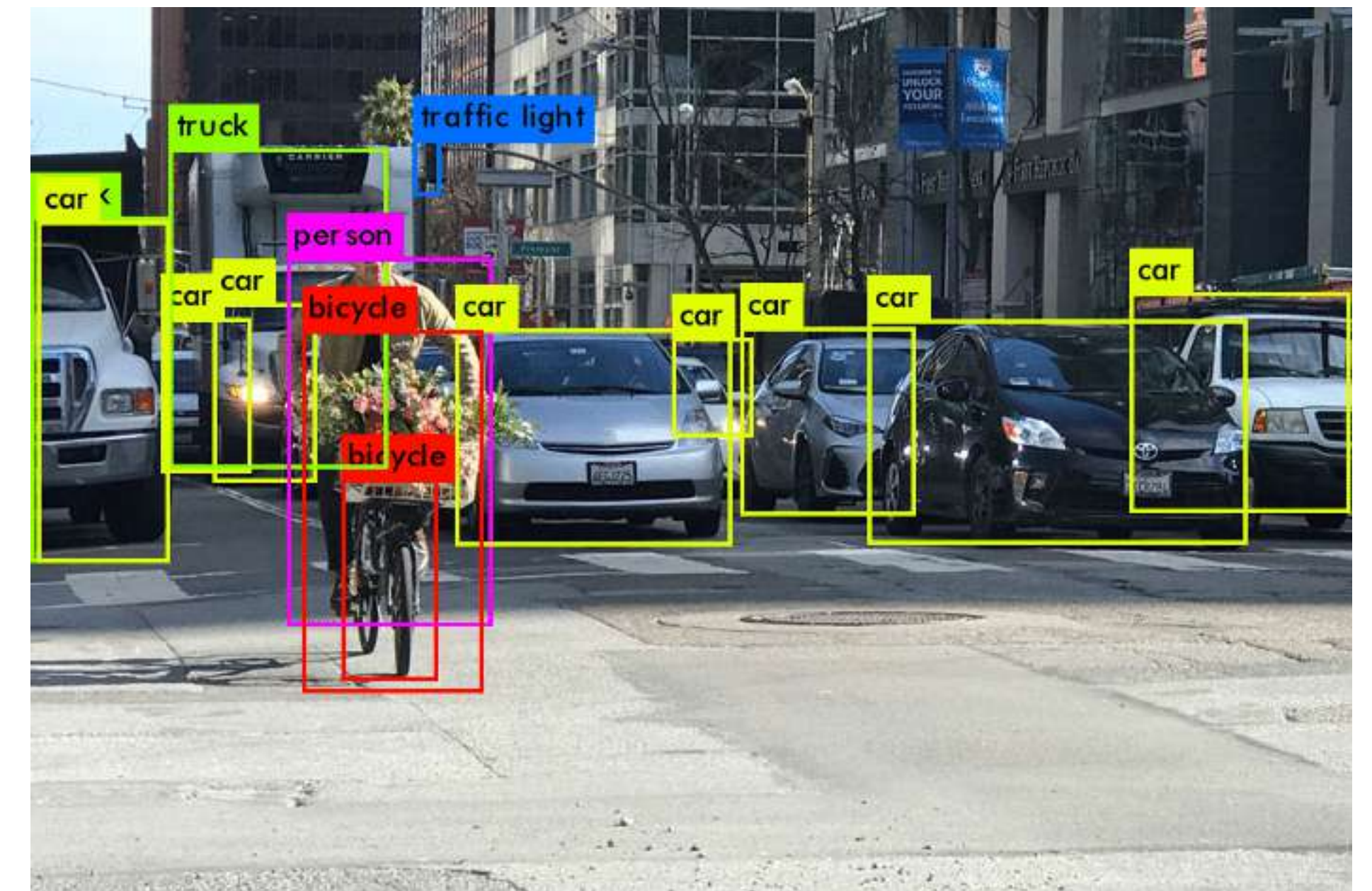


Application du machine learning

Analyse d'image



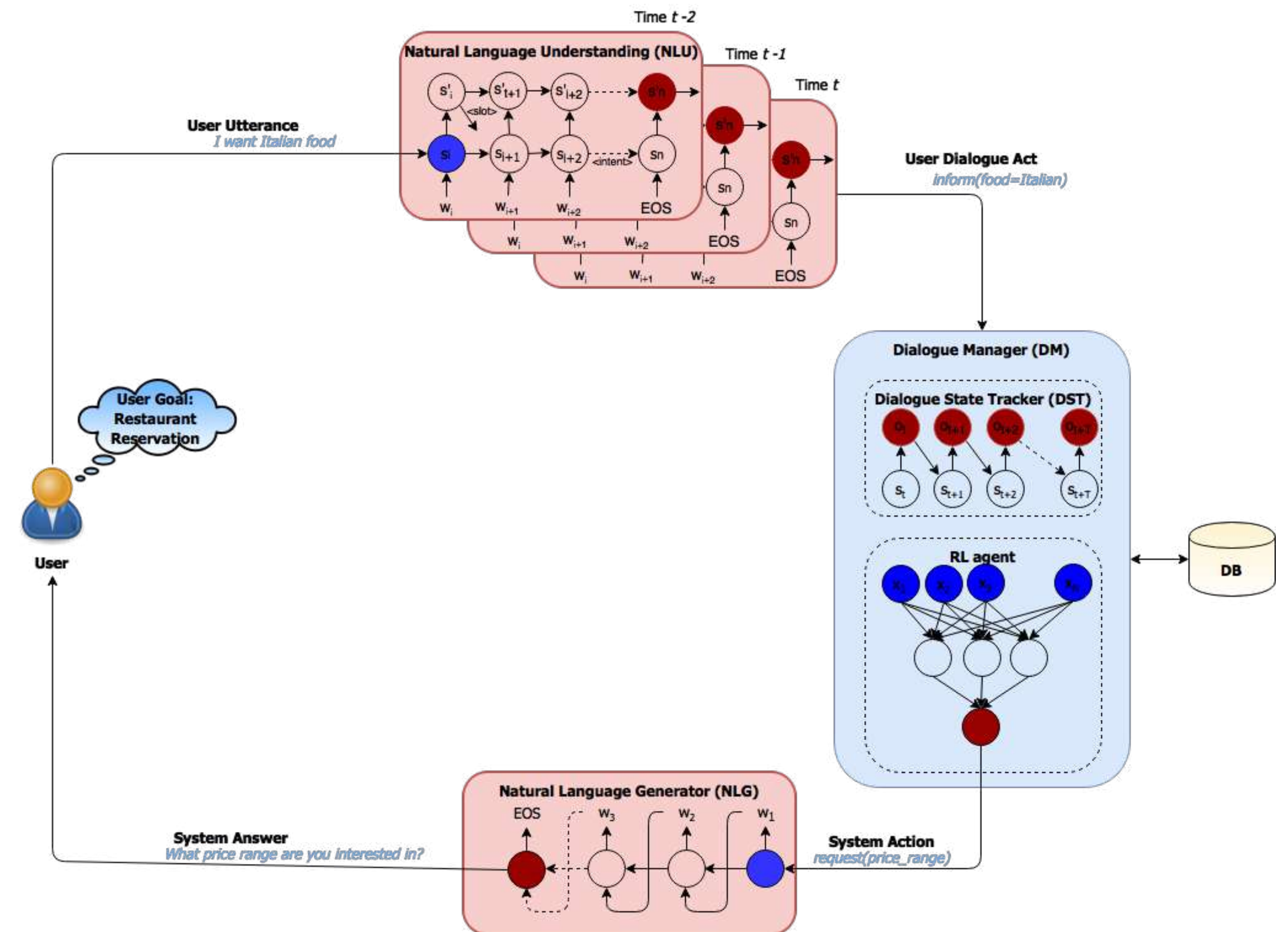
Détection et suivi en temps réel de véhicules (YOLO + tracking)




Application du machine learning

Chatbot

- Traitement et Analyse du Langage (TAL ou NLP en anglais)
- Génération de texte (chatGPT)



Machine Learning \neq Intelligence artificielle



The banner for the 36th AAAI Conference on Artificial Intelligence features a large red maple leaf logo on the left with the text 'AAAI-22' in red. To the right, a photograph shows the Vancouver skyline with the Science World dome in the foreground. Text on the banner includes '36th AAAI Conference on Artificial Intelligence', 'Vancouver, BC, Canada', and 'February 22 - March 1, 2022'. A dark blue navigation bar contains the links 'PROGRAM', 'CALLS', and 'ORGANIZATION' with dropdown arrows.

36th AAAI Conference on Artificial Intelligence
Vancouver, BC, Canada
February 22 - March 1, 2022

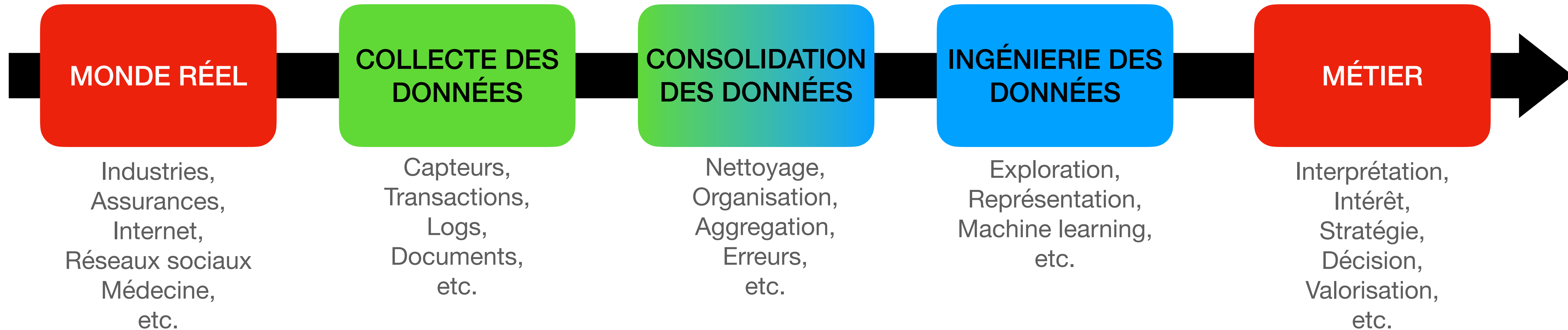
PROGRAM \vee CALLS \vee ORGANIZATION \vee

AAAI-22 Keywords

Submission Groups

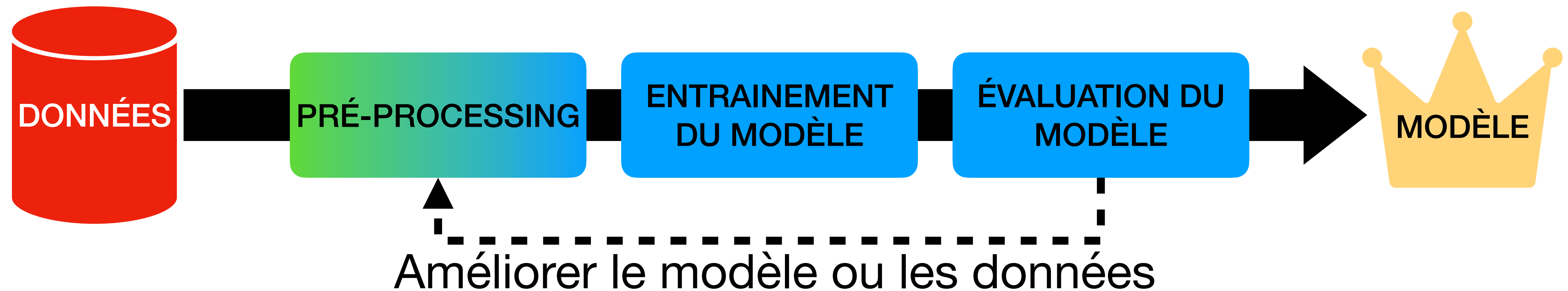
- Cognitive Modeling & Cognitive Systems (CMS)
- Computer Vision (CV)
- Constraint Satisfaction and Optimization (CSO)
- Data Mining & Knowledge Management (DMKM)
- Game Theory and Economic Paradigms (GTEP)
- Humans and AI (HAI)
- Intelligent Robotics (ROB)
- Knowledge Representation and Reasoning (KRR)
- Machine Learning (ML)
- Multiagent Systems (MAS)
- Philosophy and Ethics of AI (PEAI)
- Planning, Routing, and Scheduling (PRS)
- Reasoning under Uncertainty (RU)
- Search and Optimization (SO)
- Speech & Natural Language Processing (SNLP)
- Domain(s) of Application (APP)

Comment implémenter un projet de ML ?



Processus d'ingénierie des données

- Objectifs : obtenir un **modèle** représentatif des données pour
 - Analyser/Comprendre les données
 - Aider à la prise de décision en prédisant les données à venir



Processus d'ingénierie des données



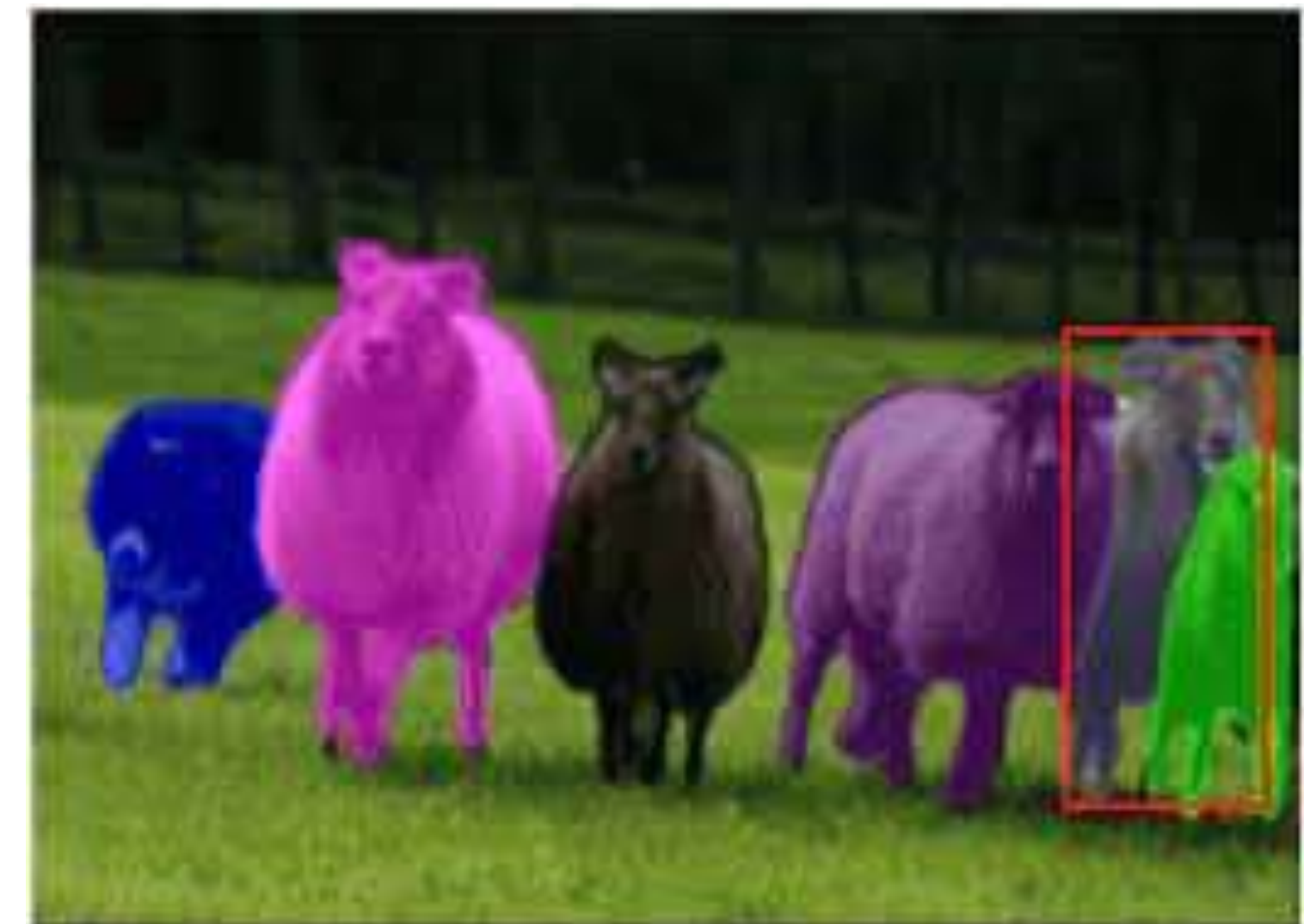
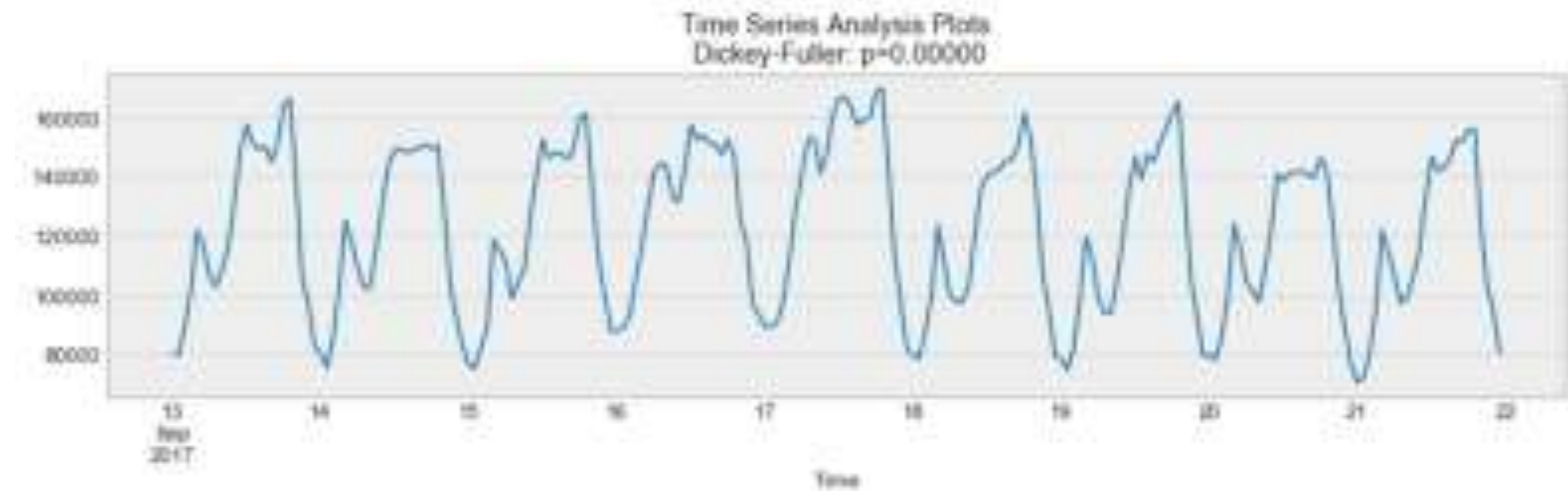
- Questions à se poser :
 - Quels sont les besoins/objectifs du client ?
 - A quel type de problème ML a-t-on affaire ?
 - Quel algorithme de ML utiliser ?
 - Quels sont les meilleurs paramètres pour cet algorithme ?
 - Comment évaluer la performance du modèle ?
 - Quel est l'impact de ce modèle pour la société ?
 - Etc.

Les données

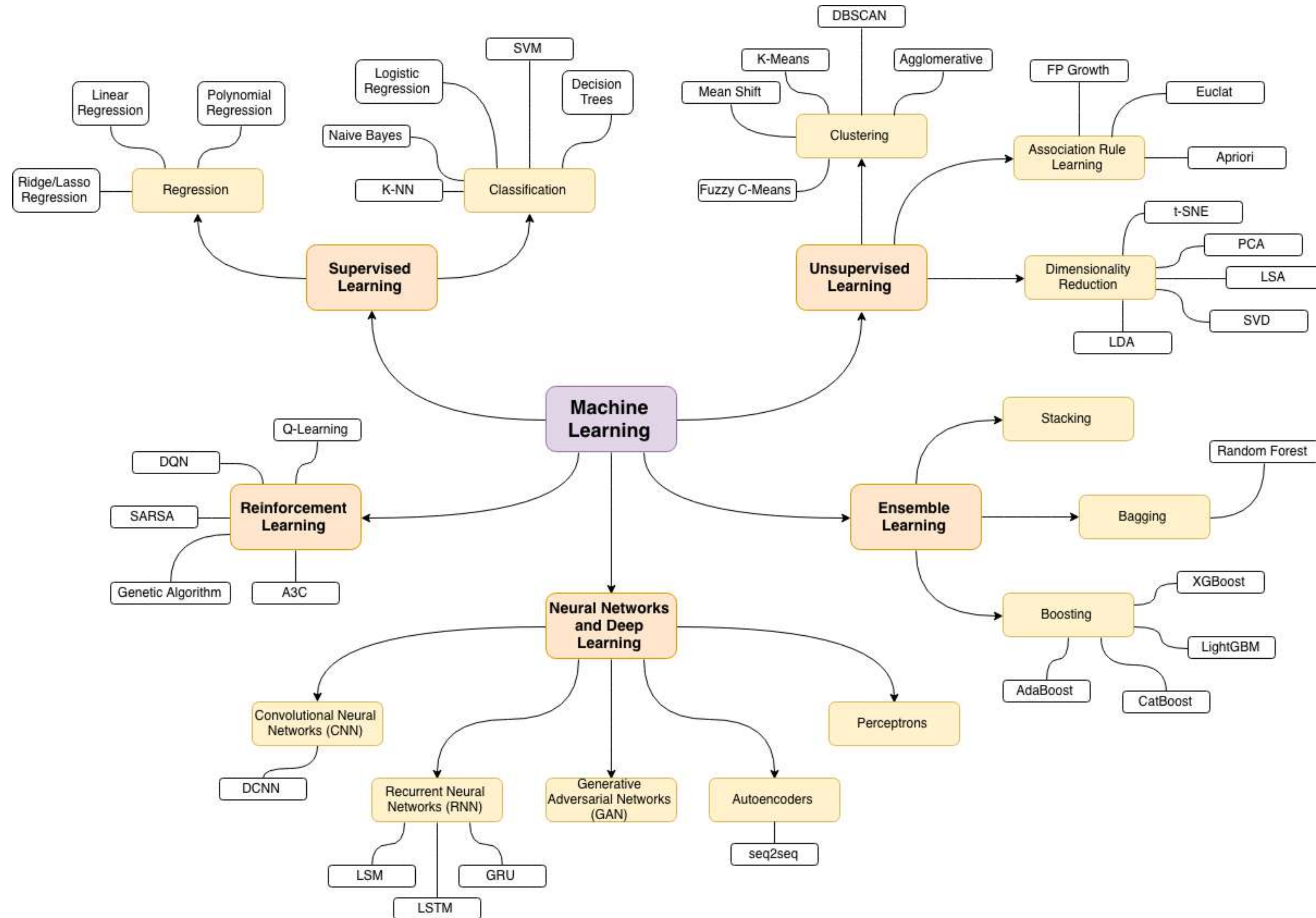
- **Ensemble de données (data set)**
 - **Exemples** d'expériences passées (informations) avec des **attributs**.
 - Un data set est constitué de N exemples (samples).
- **Attributs**
 - Un attribut est une **caractéristique (feature)** ou **propriété** d'un phénomène observé.
 - On parle aussi de variable.
- **Exemple (sample)**
 - C'est un ensemble d'attributs caractérisant un objet.
 - On parle aussi d'instance, de point ou de vecteur (souvent dans \mathbb{R}^d)

Types de données

- Capteurs :
 - Variables quantitatives et qualitatives
 - Données ordinales (échelles, intervalles, etc.)
 - Données nominales (genre, choix multiples, etc.)
- Texte : Chaîne de caractères
- Parole : Série temporelle (time series)
- Image : Donnée 2D
- Vidéo : Donnée 2D + temps
- Réseau : Graphes
- Flux : logs, tickets, etc.



Les différentes approches de ML

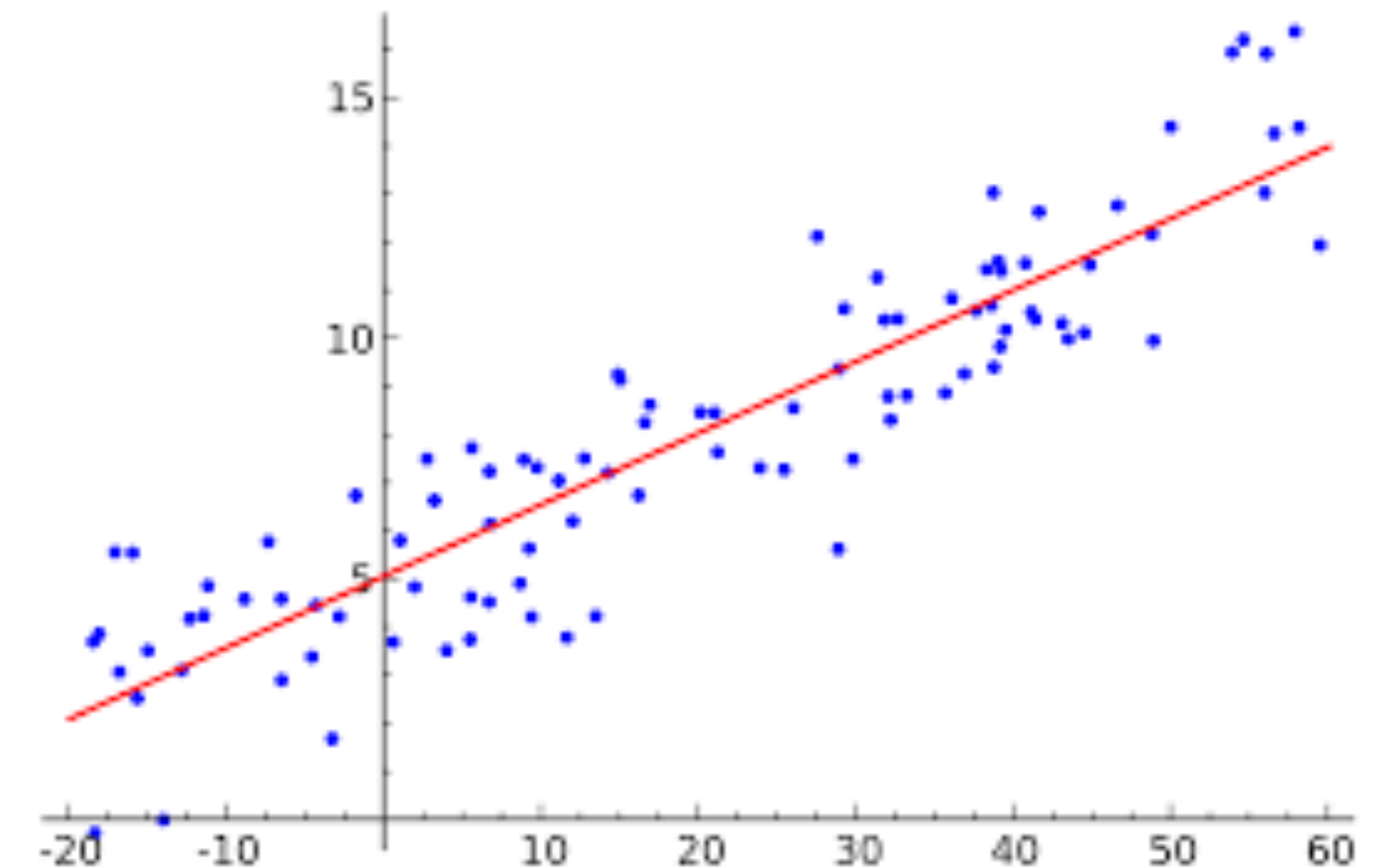


Apprentissage supervisé

Définition

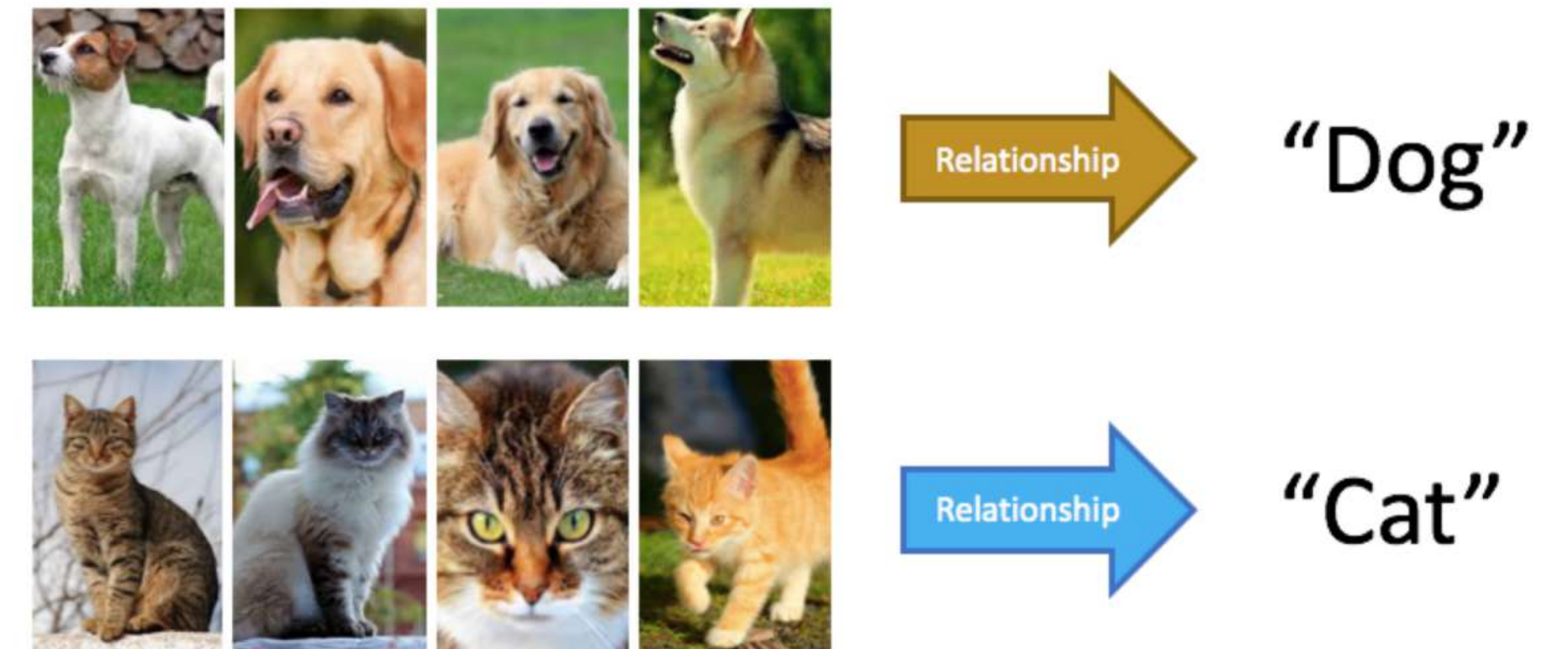
- **Principe**

- Etant donné un ensemble de N exemples d'entraînement $\{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}, i = 1..N\}$, on cherche une fonction de prédiction telle que $y = f(x)$
- **L'apprentissage est supervisé car les labels y sont connus**



- **Exemples**

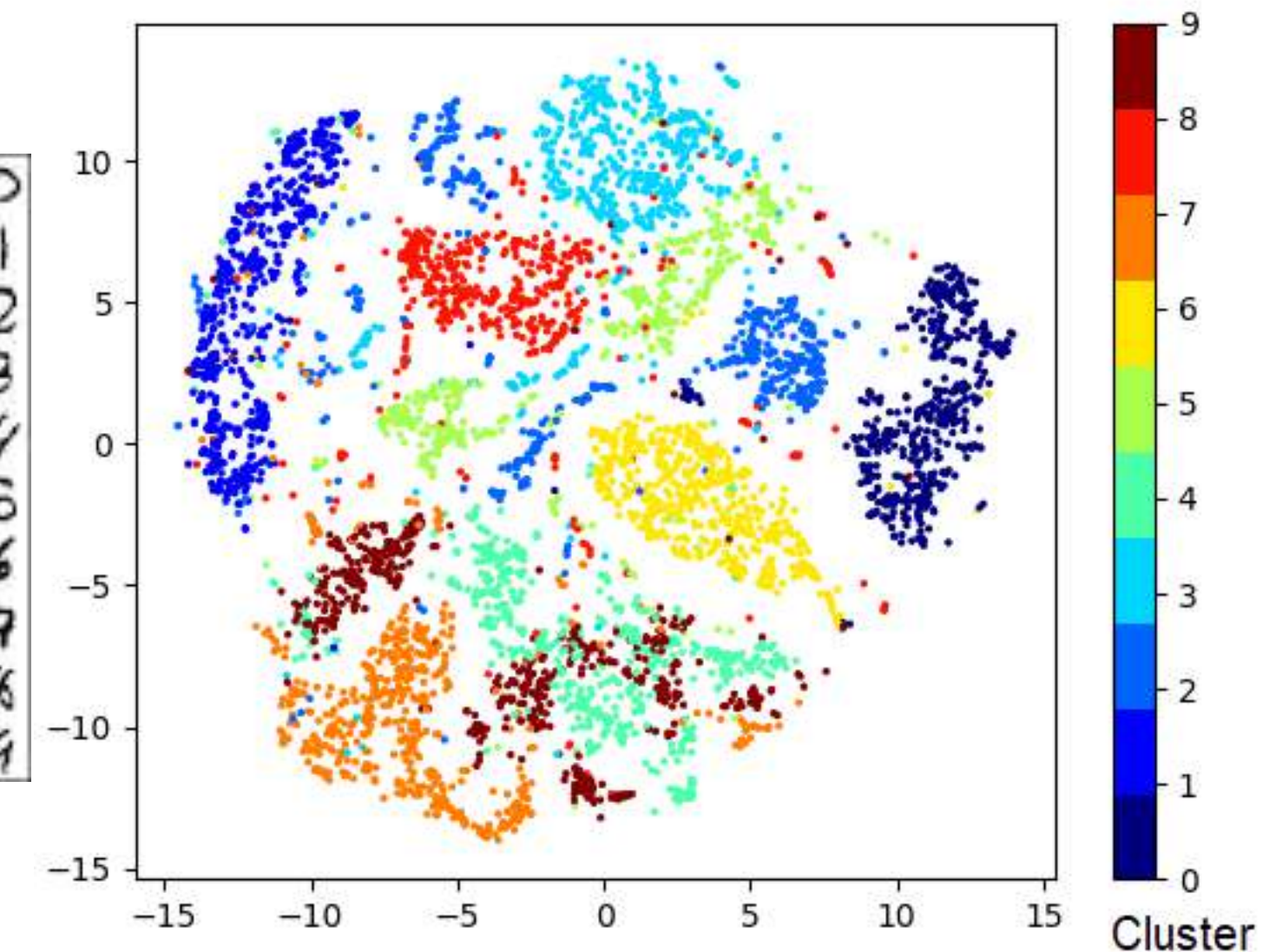
- Classification d'images
- Prédiction météo (température), stock, etc.



Apprentissage non-supervisé

Définition

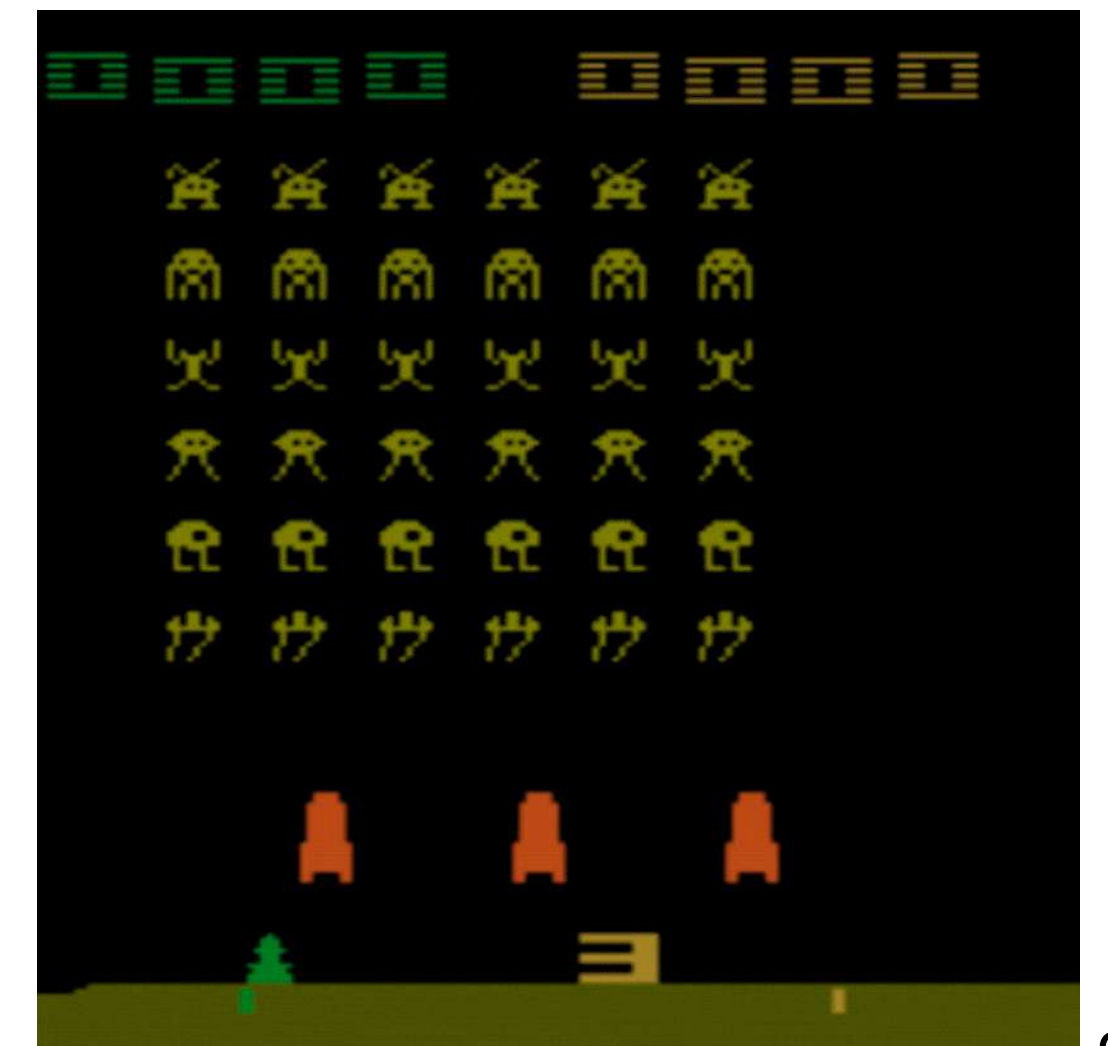
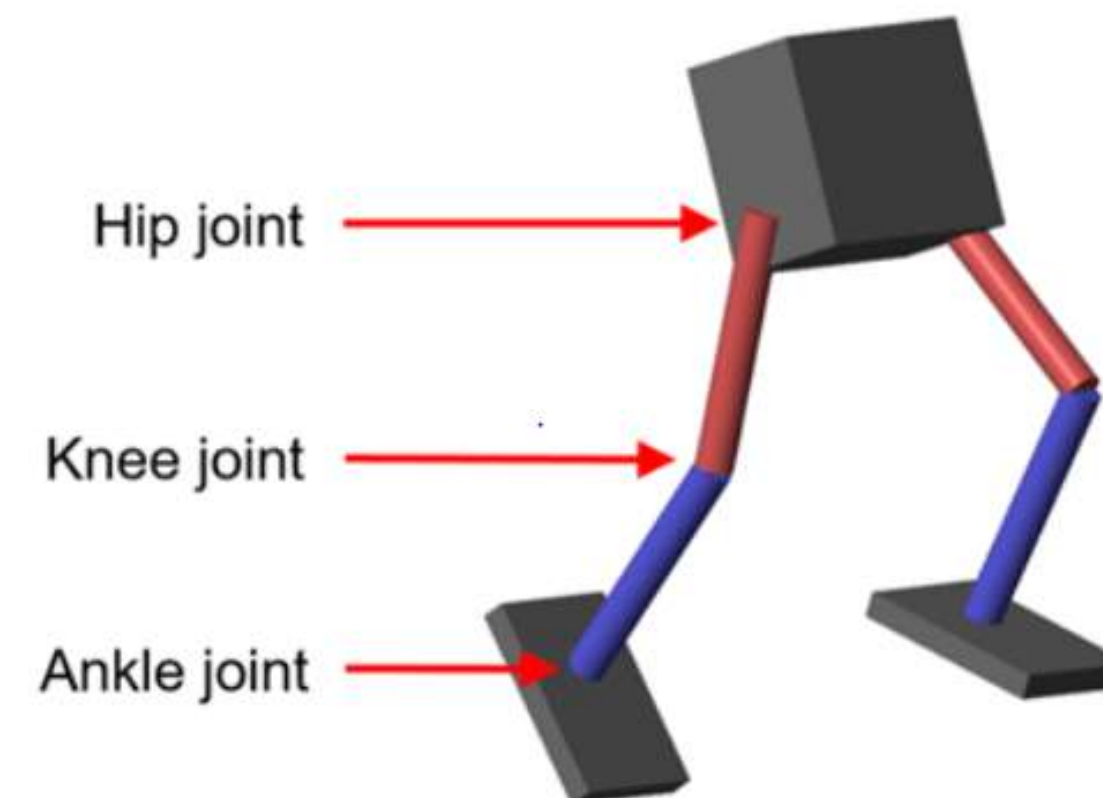
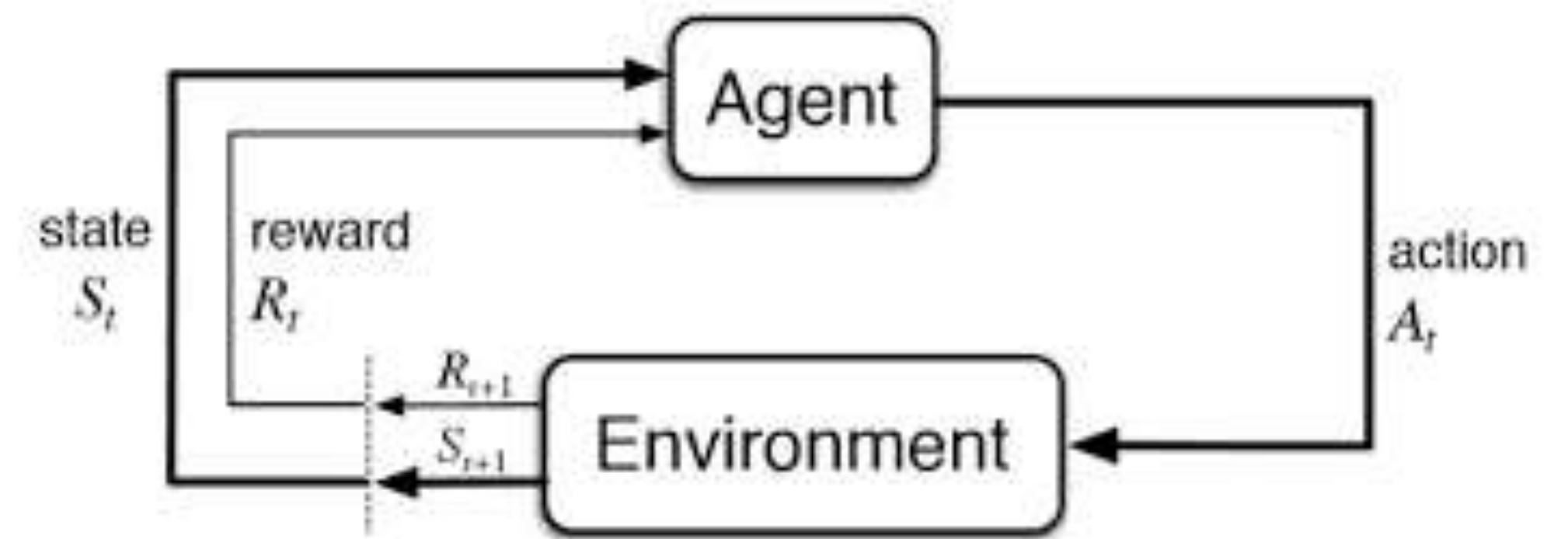
- **Principe :**
 - Seules les caractéristiques $\{x_i \in \mathcal{X}, i = 1..N\}$ sont disponibles.
 - On cherche à décrire comment les données sont **organisées** et extraire les **sous-ensemble homogènes**.
- **Exemple :**
 - Reconnaissance des formes
 - Analyse spatiales
 - Catégorisation de documents
 - Catégorisation d'usage (web, personnalisation, etc.)



Apprentissage par renforcement

Définition

- Principe
 - Pas de supervision, seulement une **récompense** (reward)
 - Le feedback n'est pas instantané
 - Le temps est crucial
 - Les actions affectent les données à venir
- Exemple
 - Jeu vidéo
 - Robotique
 - Personnalisation de contenu



- A quel type d'apprentissage avons-nous à faire pour le dataset IRIS?

[Lien vers le notebook](#)

Sources

- Quantité de données produites

X : <https://www.dsayce.com/digital-marketing/tweets-day/>

<https://medium.com/%40iam-abdulmoiz/system-design-interview-2-interviews-estimations-and-rate-limiters-91b38cb9ca84>

Facebook : <https://keywordseverywhere.com/blog/data-generated-per-day-stats/>

- IoT :

<https://www.sap.com/france/products/technology-platform/what-is-iot.html#:~:text=Histoire%20de%20l'IoT,73%20000%20milliards%20de%20gigaoctets.>

-