Faculté des sciences de la terre et de l'architecture

Département Géographie et aménagement du territoire

Chapitre 3

L'APPRENTISSAGE NON SUPERVISÉ

QU'EST-CE QUE L'APPRENTISSAGE NON SUPERVISÉ ?

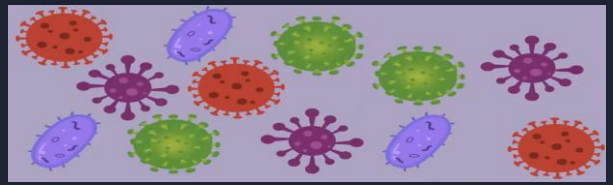
Définition

Contrairement à l'approche supervisée, l'apprentissage non supervisé traite des données qui ne possèdent pas d'étiquettes ou de "réponses" prédéfinies.

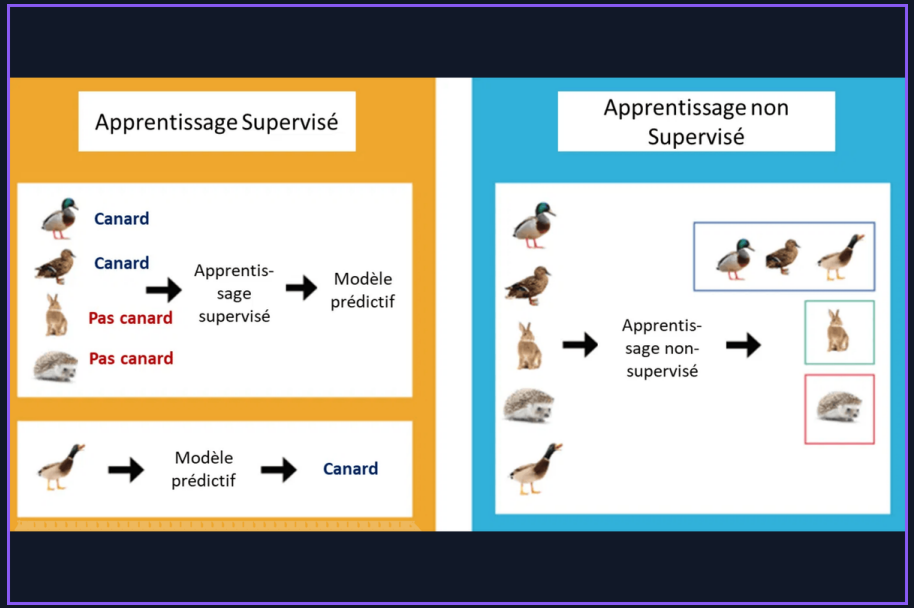
La machine reçoit un ensemble d'informations brutes et doit, par elle-même, identifier des **motifs**, des **régularités** ou des **regroupements naturels** au sein de ces données.

Imaginez un chercheur observant au microscope une multitude de virus et de bactéries de formes variées, sans savoir à quelle maladie ils correspondent.

L'algorithme va trier ces agents pathogènes en fonction de leurs ressemblances physiques (forme, structure), créant ainsi des familles distinctes simplement en observant leurs caractéristiques communes.



SUPERVISÉ VS NON SUPERVISÉ



Apprentissage Supervisé

On donne à la machine des exemples avec des **étiquettes** (Canard / Pas canard). Elle apprend à reconnaître le canard grâce à ces réponses fournies par l'expert.

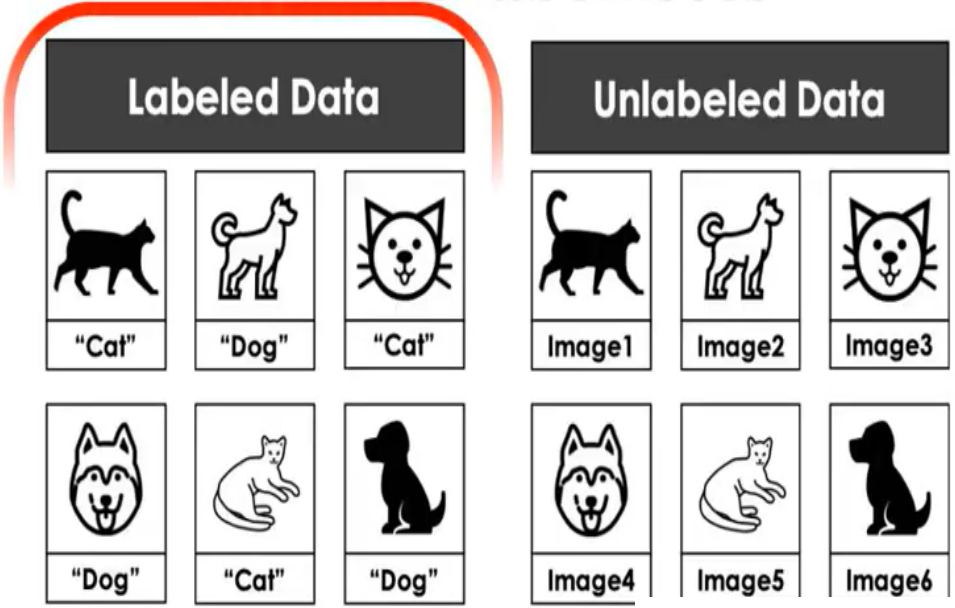
Apprentissage Non Supervisé

On donne toutes les photos à la machine sans rien lui dire. Elle va naturellement **regrouper** les photos qui se ressemblent (tous les canards ensemble, tous les lapins ensemble) car ils partagent des caractéristiques communes.

Dataset

La machine connaît déjà les réponses qu'on attend d'elle. Elle travaille à partir de

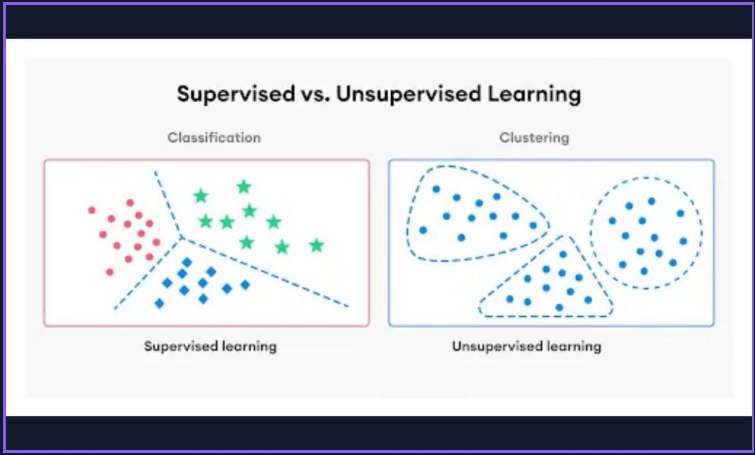
données labellisées.



Les réponses que l'on cherche à prédire ne sont pas disponibles dans les jeux de données.

SUPERVISÉ VS NON SUPERVISÉ

Caractéristique	Supervisé	Non Supervisé
Données	Étiquetées	Non étiquetées
Objectif	Prédire	Découvrir
Exemple	Spam / Non Spam	Segmentation
Expert	Donne les étiquettes	Définit la similarité



LES MISSIONS DE L'APPRENTISSAGE NON SUPERVISÉ

Le Clustering

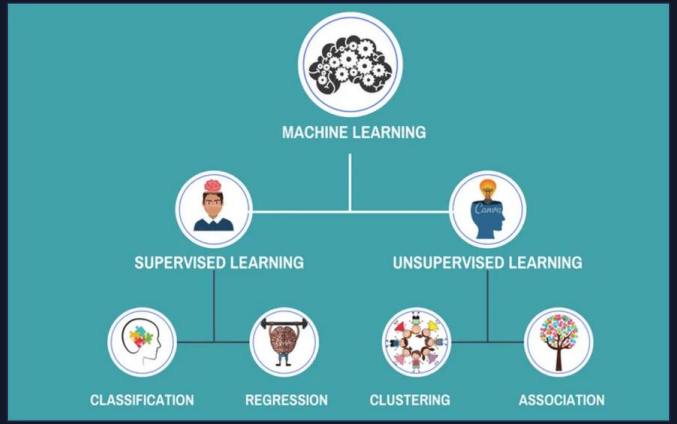
Diviser un ensemble de données en sous-groupes homogènes. Les éléments d'un même groupe partagent des caractéristiques communes et diffèrent des autres groupes.

Réduction de Dimension

Simplifier des données volumineuses en ne conservant que les informations les plus pertinentes. Crucial pour la visualisation et l'efficacité des calculs.

L'Association

Découvrir des règles de dépendance entre les variables. Par exemple, identifier que l'achat d'un produit spécifique en entraîne souvent un autre.



CAS D'ÉTUDE : AMÉNAGEMENT URBAIN

La Problématique

Comment diviser une métropole en **zones homogènes** pour optimiser le déploiement des services publics et des infrastructures ?

Données Utilisées

- Densité de population par quartier
- Revenus moyens des ménages
- Proximité des réseaux de transport

Résultats du Clustering

L'algorithme identifie naturellement des quartiers aux profils distincts : **résidentiels, commerciaux ou industriels**.

Cette analyse révèle des dynamiques urbaines sans avoir besoin de définir manuellement chaque zone à l'avance.

Impact Décisionnel

Permet aux urbanistes de cibler les investissements et de mieux comprendre l'évolution de la structure urbaine.

LES AVANTAGES

Autonomie et Économie

Permet de traiter d'immenses volumes de données sans nécessiter l'intervention coûteuse d'experts pour l'étiquetage manuel.

Découverte de l'Inconnu

Capable de révéler des structures et des corrélations que l'œil humain ou des analyses statistiques classiques pourraient ignorer.

EXERCICE

01

Traitement des Données Massives

Expliquez pourquoi l'apprentissage non supervisé est particulièrement adapté au traitement des données massives (**Big Data**) par rapport à l'approche supervisée.

02

Clustering vs Classification

Quelle est la différence fondamentale entre le "Clustering" et la "Classification"?

SOLUTION

- **Absence d'étiquetage manuel** : Dans le Big Data, nous avons des millions de données. Il serait impossible et extrêmement coûteux de demander à des humains de toutes les étiqueter à la main.
- **Découverte automatique** : La machine explore seule les données pour trouver des structures cachées que l'humain ne pourrait pas voir dans une telle masse d'informations.
- **Rapidité de mise en œuvre** : On peut lancer l'algorithme directement sur les données brutes pour obtenir une première segmentation rapide sans préparation préalable complexe.

L'apprentissage non supervisé est l'outil idéal pour donner du sens à des données massives et désorganisées.

CLASSIFICATION (SUPERVISÉ)

On connaît déjà les classes (ex: "Maison A", "Maison B"). La machine apprend à classer de nouvelles données dans ces **catégories connues** grâce aux étiquettes fournies par l'expert.

CLUSTERING (NON SUPERVISÉ)

On ne connaît pas les classes. La machine **découvre elle-même** des groupes naturels en fonction de la ressemblance des données, sans aucune étiquette préalable.

QUIZ

1. Dans l'apprentissage non supervisé, les données sont :

A) Étiquetées par un expert

B) Non étiquetées

C) Accompagnées de la réponse

2. Le Clustering sert à :

A) Prédire le prix d'une maison

B) Regrouper des objets similaires

C) Traduire un texte

QUIZ : VRAI OU FAUX ?

1. L'apprentissage non supervisé est plus rapide à préparer car il ne nécessite pas d'étiquetage manuel. **VRAI**

2. Dans l'apprentissage non supervisé, la machine connaît déjà le résultat final attendu. **FAUX**

3. L'aménagement urbain peut utiliser le clustering pour identifier des quartiers similaires. **VRAI**

4. Le clustering est une méthode d’apprentissage supervisé. **FAUX**

5. Le clustering peut regrouper des clients ayant des comportements similaires. **VRAI**