

Faculté des sciences de la terre et de l'architecture

Département Géographie et aménagement du territoire

# Intelligence Artificielle

---

Apprentissage Non Supervisé – Clustering

## DBSCAN

**Bezzaz Soumia**

2025/2026 – Semestre 2

# Qu'est-ce que DBSCAN ?

DBSCAN = Density-Based Spatial Clustering of Applications with Noise

## Définition simple

DBSCAN est un algorithme non supervisé qui trouve des **groupes denses** de points dans l'espace.

Contrairement à K-means :

- Ne nécessite pas de fixer K à l'avance
- Détecte les formes arbitraires
- Identifie automatiquement le **bruit** (outliers)

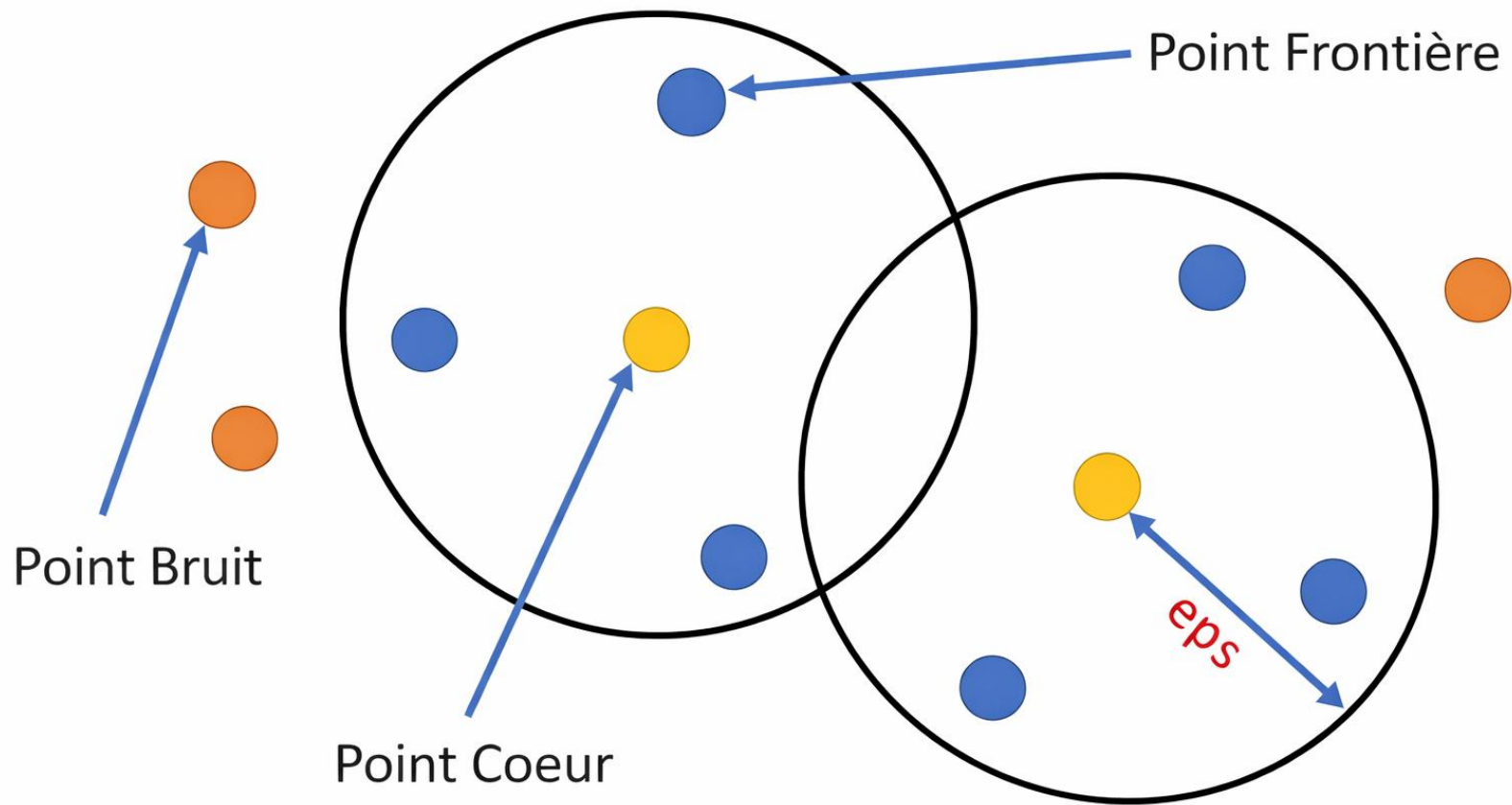
## Concepts clés

### Basé sur la densité

Les clusters sont des régions où les points sont concentrés.

*Deux paramètres principaux :*

- **Epsilon ( $\epsilon$ )** : rayon de voisinage
- **MinPts** : nombre minimum de points



Point Frontière

Point Bruit

Point Coeur

minPoints = 3

# Analogie : Identifier les zones forestières denses

Imaginez une région avec des arbres dispersés et des zones de forêt dense



## Le problème

- Vous avez une carte satellite d'une région
- Chaque pixel = présence/absence d'arbres
- Objectif : identifier les **zones de forêt dense** (clusters) vs arbres isolés (bruit)
- Comment définir une "forêt" ?

## La solution DBSCAN

1. Fixez  $\epsilon = 50\text{m}$  (rayon de recherche)
2. Fixez **MinPts = 10** arbres minimum
3. DBSCAN scanne chaque arbre :
  - S'il a  $\geq 10$  voisins dans 50m = **cœur de forêt**
  - Sinon = arbre isolé (bruit)
4. Les arbres-cœur forment des clusters

✓ *Détecte formes irrégulières de forêts*

# Comment DBSCAN fonctionne : Les 4 étapes

1

## CLASSIFICATION

### Classifier chaque point

Pour chaque point, compter combien de voisins sont dans un rayon  $\epsilon$  :

- Si  $\geq \text{MinPts}$  → **Point-cœur**
- Sinon → Point-bordure ou Bruit

2

## CONNEXION

### Connecter les points-cœur

Si deux points-cœur sont à distance  $\leq \epsilon$ , ils appartiennent au même cluster.

Créer des chaînes de points-cœur connectés.

3

## ASSIGNATION

### Assigner les points-bordure

Les points qui ne sont pas des cœurs mais qui sont dans le voisinage  $\epsilon$  d'un point-cœur rejoignent ce cluster.

4

## BRUIT

### Identifier le bruit (outliers)

Les points qui ne sont ni cœurs ni bordures sont marqués comme **bruit** (anomalies).

# Les paramètres : $\epsilon$ (epsilon) et MinPts

## $\epsilon$ (Epsilon)

*Distance maximale de voisinage*

### Définition :

Le rayon autour d'un point pour chercher ses voisins.

### Comment choisir $\epsilon$ ?

- Trop petit  $\rightarrow$  beaucoup de bruit
- Trop grand  $\rightarrow$  tous les points dans 1 cluster

### Méthode : *Graphique k-distance*

Calculer la distance au k-ème voisin le plus proche pour tous les points, puis chercher le "coude" dans la courbe.

## MinPts

*Nombre minimum de points*

### Définition :

Le nombre minimum de points dans le voisinage  $\epsilon$  pour qu'un point soit considéré comme point-cœur.

### Comment choisir MinPts ?

- Règle empirique :  $\text{MinPts} \geq \text{dimensions} + 1$
- Pour 2D :  $\text{MinPts} = 4$  est courant
- Plus MinPts est élevé  $\rightarrow$  clusters plus denses requis

# DBSCAN vs K-means : Comparaison

Critère	K-means	DBSCAN
Nombre de clusters	K fixé à l'avance	Découvert automatiquement
Forme des clusters	Sphériques uniquement	Formes arbitraires
Gestion du bruit	Tous points assignés	Détecte et isole le bruit
Taille des clusters	Tailles similaires	Tailles variables OK
Paramètres	K (nombre de clusters)	$\epsilon$ (epsilon) et MinPts
Complexité	rapide	plus lent
Données spatiales	Moins adapté	Excellent

# Applications de DBSCAN dans le monde réel



## Géosciences & SIG

- Détection de zones urbaines denses
- Identification de hotspots sismiques
- Délimitation de bassins versants
- Analyse de pollution spatiale

💡 *Cartographie automatique des zones à risque d'inondation*



## Climatologie & Météo

- Détection de zones climatiques
- Identification de cyclones/tornades
- Analyse de patterns météo
- Délimitation de microclimats

💡 *Prédiction de zones de sécheresse à partir de données satellites*



## Analyse urbaine

- Segmentation de quartiers
- Analyse de trafic routier
- Détection de zones commerciales
- Planification territoriale

💡 *Optimisation de l'emplacement de services publics*



## Science des données

- Détection d'anomalies réseau
- Analyse de comportements clients
- Segmentation de marchés
- Identification de fraudes

💡 *Détection de transactions frauduleuses en temps réel*

# Cas pratique : Segmentation spatiale de l'occupation des sols

Exemple concret : détecter automatiquement les zones d'occupation similaires

## Le défi

### Données :

Image Sentinel-2 d'une région agricole (20km × 20km).

Chaque pixel contient :

- Indice de végétation NDVI
- Humidité du sol
- Altitude

### Objectif :

Identifier les zones géographiquement cohérentes (forêt, culture, urbain, eau) sans connaître à l'avance le nombre de types.

## Solution avec DBSCAN

### Paramètres choisis :

- $\epsilon = 100\text{m}$  (pixels voisins)
- MinPts = 20 pixels

### Résultat :

DBSCAN identifie automatiquement :

- ✓ 3 clusters de forêts denses
- ✓ 5 zones agricoles
- ✓ 2 zones urbaines
- ✓ 1 rivière (forme sinueuse)
- ✓ Pixels de transition = bruit

*Temps : ~5 minutes pour 4 millions de pixels*

# Exercice 1 :

## Contexte :

Un géographe analyse 10 stations météo et mesure leur température moyenne annuelle (en °C).  
Il veut regrouper les stations en zones climatiques similaires en utilisant DBSCAN avec  $\epsilon = 3^\circ\text{C}$  et  $\text{MinPts} = 3$ .

Station	Reykjavik	Oslo	Londres	Paris	Lisbonne
Temp (°C)	5°C	6°C	11°C	12°C	17°C
Station	Athènes	Le Caire	Dubaï	Mumbai	Djibouti
Temp (°C)	19°C	22°C	28°C	27°C	30°C

## Questions :

- Q1. Quelles stations seront classées comme points-cœur ?
- Q2. Combien de clusters DBSCAN va-t-il former ?
- Q3. Y aura-t-il des points considérés comme bruit ? Si oui, lesquels ?

# Exercice 2 : Vrai ou Faux

DBSCAN nécessite de fixer le nombre de clusters  $K$  à l'avance.

V F

→ F

DBSCAN peut détecter des clusters de formes arbitraires (non-sphériques).

V F

→ V

Tous les points sont assignés à un cluster dans DBSCAN.

V F

→ F

Les paramètres  $\epsilon$  et MinPts doivent être choisis par l'utilisateur.

V F

→ V

DBSCAN est plus rapide que K-means sur de grandes données.

V F

→ F

Un point-cœur a au moins MinPts voisins dans son rayon  $\epsilon$ .

V F

→ V

# À retenir sur DBSCAN

---

1

## Algorithme basé sur la densité

Groupe les points denses et isole le bruit automatiquement.

2

## Deux paramètres clés

$\epsilon$  (rayon) et MinPts (seuil de densité) définissent les clusters.

3

## Formes arbitraires

Détecte des clusters non-sphériques, idéal pour données spatiales.

4

## Pas besoin de K

Découvre le nombre de clusters automatiquement.

5

## Gestion du bruit

Identifie les outliers (anomalies) contrairement à K-means.

6

## Applications variées

Géosciences, climatologie, urbanisme, détection d'anomalies.