

# Clustering (DBSCAN)

## تقنية التجميع (Clustering):

هي عبارة عن عملية تقسيم مجموعة من البيانات غير المصنفة إلى مجموعات فرعية أو مجموعات تجميع بناءً على الشبهات الداخلية بين البيانات. يتم تجميع البيانات في نفس المجموعة إذا كانت تتشابه بشكل كبير في بعض الخصائص أو السمات، في حين تكون المجموعات المختلفة متميزة عن بعضها البعض.

### DBSCAN:

هو اختصار لـ Density-Based Spatial Clustering of Applications with Noise، وهو خوارزمية تجميع (clusters) بيانات قائمة على الكثافة المكانية مع التعامل مع النقاط الفردية كـ (noise). وتستند هذه الخوارزمية إلى فكرة أن المجموعات (clusters) في البيانات تتكون من مناطق عالية الكثافة، وأنها مفصولة بواسطة مناطق منخفضة الكثافة. يهدف DBSCAN إلى اكتشاف وتجميع المناطق عالية الكثافة وتجاوز المناطق المنخفضة الكثافة.

بالإضافة إلى ذلك، يتميز DBSCAN بأنه لا يتطلب معرفة مسبقة لعدد التجمعات، مما يعني أنه يمكنه اكتشاف عدد التجمعات تلقائيًا بناءً على الكثافة وهيكل البيانات.

### تستند خوارزمية DBSCAN إلى مفهومين رئيسيين:

- النقاط الأساسية (core points): هي النقاط التي لديها عدد من الجيرة الأكبر من قيمة معينة (تحدد مسبقًا) في نطاق محدد.
  - والجيرة (neighborhood): تشمل النقاط التي تقع ضمن نطاق محدد من نقطة معينة.
- بشكل عام، يتم تطبيق DBSCAN في مجالات مختلفة مثل تحليل البيانات، وعلوم البيانات، والتعلم الآلي، والرؤية الحاسوبية، وغيرها من المجالات التي تتطلب تجميع البيانات المكانية بناءً على الكثافة.

**فكرة الخوارزمية:** يقوم DBSCAN بتجميع النقاط المتقاربة في مجموعات ويتجاوز النقاط المنفردة أو الضوضاء.

## آلية عمل هذه الخوارزمية :

نقوم بتحديد التالي:

1. نقطة أساسية (Core Point): نبدأ بنقطة ونحدد نطاقًا حولها. إذا كانت هناك نقاط كثيرة داخل هذا النطاق (أكبر من عدد معين مسبقًا)، فإن النقطة تعتبر نقطة أساسية.
2. الجيرة (Neighborhood): نحدد نطاقًا حول كل نقطة أساسية. يتم اعتبار النقاط التي تقع ضمن هذا النطاق كجيرة.
3. توسيع التجمع (Cluster Expansion): يتم تمديد التجمع بإضافة النقاط المتصلة بالنقطة الأساسية. إذا كانت النقاط المتصلة أيضًا نقاط أساسية، يتم تمديد التجمع بالنقاط المتصلة بها أيضًا، وهكذا.
4. الضوضاء (Noise): النقاط التي ليس لها نقاط أساسية في نطاقها تعتبر ضوضاء ولا تنتمي إلى أي تجمع.

### مثال :

لنفترض أن لدينا مجموعة من البيانات ونرغب في تجميعها باستخدام DBSCAN.