

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algoritması, yoğunluk temelli bir kümeleme yöntemidir. Yani yoğun veri noktalarını kümeler ve seyrek alanlardaki noktaları **aykırı (outlier)** olarak işaretler.

Min_samples Nedir?

Bir noktanın "**çekirdek nokta**" (**core point**) olabilmesi için çevresinde en az min_samples kadar **yakın komşusu** olmalıdır (bu komşuların uzaklığı eps'ten küçük olmalı).

min_samples = minimum yoğunluk kriteri.

min_samples Ne İşe Yarar?

Düşük min_samples:

- Daha **fazla küçük küme** çıkarabilir.
- Gürültülü verileri bile kümeye alabilir.
- Aykırı değer sayısı az olur.

Yüksek min_samples:

- Sadece **çok yoğun bölgeleri** küme olarak kabul eder.
- Daha fazla **aykırı değer (outlier)** çıkarabilir.
- Belki bazı gerçek kümeleri bile aykırı gibi görebilir.

min_samples Nasıl Seçilir?

Pratik Kural:

- **min_samples = veri boyutu / 100**
- Ya da:
 - **2D veri** → en az **4**
 - Daha yüksek boyutlu verilerde (n_features), genellikle:
 $\text{min_samples} \geq \text{n_features} + 1$

Yani bu örnekte:

```
X = df[["total_orders", "total_spent", "avg_order_value"]]
```

```
# n_features = 3 → min_samples en az 4 olabilir
```

Özetle;

Parametre	Ne işe yarar?
eps	Komşuluk yarıçapı – bir noktanın çevresi
min_samples	Bir noktanın çekirdek olması için gereken komşu sayısı
Düşük değer	Küçük kümeler, az aykırı değer
Yüksek değer	Daha sıkı kümeler, çok aykırı değer