DBSCAN Clustering

DBSCAN stands for **D**ensity-**B**ased **S**patial **C**lustering of **A**pplications with **N**oise.

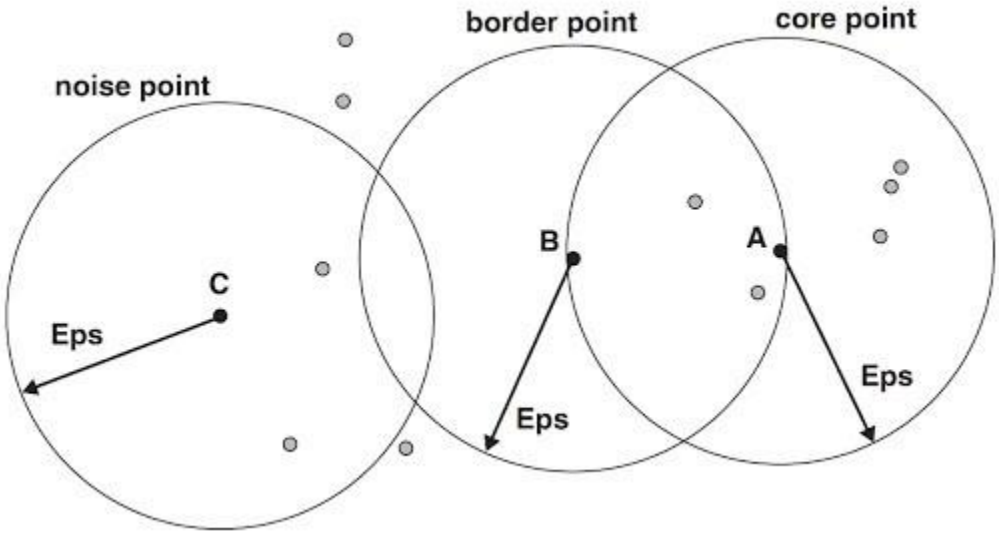
It was proposed by Martin Ester et al. in 1996. DBSCAN Algorithm is a density-based clustering algorithm that works on the assumption that clusters are dense regions in space separated by regions of lower density.

It groups 'densely grouped' data points into a single cluster. It can identify clusters in large spatial datasets by looking at the local density of the data points. The most exciting feature of DBSCAN clustering is that it is robust to outliers. Unlike K-Means, where we have to specify the number of centroids, it also does not require the number of clusters to be told beforehand.

The DBSCAN algorithm requires only two parameters:

- epsilon and
- minPoints.

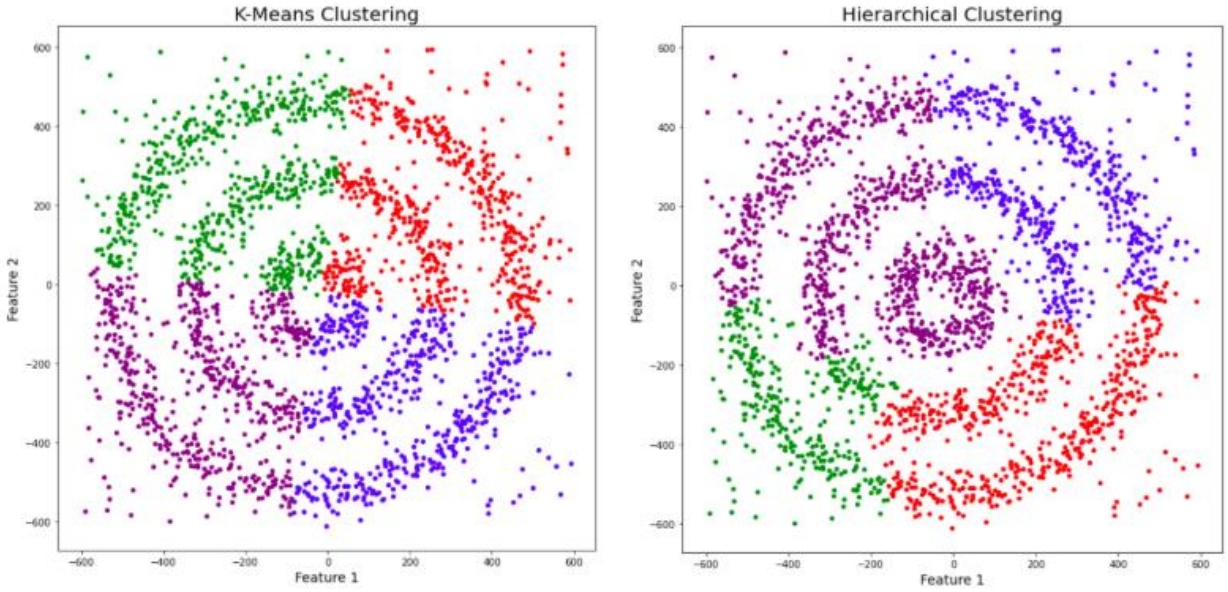
Epsilon is the radius of the circle to be created around each data point to check its density. ***minPoints*** is the minimum number of data points required inside that circle for that data point to be classified as a **Core** point.



K-Means VS DBSCAN Clustering

K-Means and Hierarchical Clustering both fail to create clusters of arbitrary shapes. They are not able to form clusters based on varying densities. That's why we need DBSCAN clustering.

We can see three different dense clusters in the form of concentric circles with some noise here. Now, let's run K-Means and Hierarchical clustering algorithms and see how they cluster these data points.



You might be wondering why there are four colors in the graph. As I said earlier, this data contains noise, too. Therefore, I have taken noise as a different cluster, which is represented by the purple color. Sadly, both of them failed to cluster the data points. Also, they were not able to detect the noise present in the dataset properly. Now, let's take a look at the results from DBSCAN clustering.