K-MEANS CLUSTERING: HOW IT WORKS AND WHY SUITABLE FOR IRIS DATASET

K-Means clustering is an unsupervised machine learning algorithm used to partition a dataset into a set number of clusters (k). The algorithm follows these steps:

- Randomly selects `k` initial centroids, where `k` is the number of desired clusters.
- Each data point is assigned to the nearest centroid based on the Euclidean distance, forming `k` clusters.
- The centroid of each cluster is recalculated as the mean of all data points in the cluster.
- The assignment and update steps are repeated until the centroids no longer change significantly or a specified number of iterations is reached.
- The algorithm converges when the cluster assignments become stable, meaning data points no longer switch between clusters.

The Iris dataset has three known species of iris flowers, which are inherently distinct groups. K-Means is well-suited for datasets where clusters can be distinguished based on distance, making it ideal for this case.The Iris dataset has only four features (sepal length, sepal width, petal length, and petal width), making it manageable for K-Means, which performs well on low to medium-dimensional data.K-Means works best when clusters are spherical and well-separated, which roughly aligns with the characteristics of the Iris dataset, where the clusters of different species tend to be distinct.

HIERARCHIAL CLUSTERING: HOW IT WORKS AND WHY SUITABLE FOR IRIS DATASET

Hierarchical clustering is an unsupervised machine learning technique that builds a hierarchy of clusters. It can be divided into two main types:

- Starts with each data point as its own cluster.
- In each step, the two closest clusters are merged.
- This process continues until all data points belong to a single cluster or a specified number of clusters is reached.

Hierarchical clustering does not require specifying the number of clusters in advance. The dendrogram can help visualize the structure and determine the optimal number of clusters for the Iris dataset. The Iris dataset contains three species, and hierarchical clustering's dendrogram can reveal relationships and sub-cluster patterns that may not be apparent with other clustering methods. Hierarchical clustering can be computationally expensive for large datasets, but it works well on smaller datasets like Iris, which contains only 150 samples.If the clusters in the Iris dataset are not perfectly spherical or well-separated, hierarchical clustering may better capture the underlying structure than K-Means, which assumes spherical clusters.

Hierarchical clustering provides a way to explore different levels of clustering in the Iris dataset, making it suitable for understanding patterns and relationships among the species.