**Clustering Analysis Using K-Means Algorithm**

K-Means is an iterative algorithm that partitions a dataset into K-predefined clusters, with each data point belonging to a single cluster. It aims to minimize the squared distance between data points and their cluster's centroid, ensuring that intra-cluster points are as similar as possible while keeping clusters distinct. Reduced variation within clusters results in greater data point homogeneity within each cluster. The approach K-means undergoes is called Expectation Maximation. This E-step assigns the data points to the nearest cluster while the M-step computes the centroid of every cluster of the data set. In our study, we assume that we don’t have a target to predict, hence it is an unsupervised learning problem [clustering analysis]. Clustering algorithms that include k-means, we used distance-based measurement to analyze the similarity of different data points, but it is recommended to standardize the data to have a mean of zero and a standard deviation of one, since the features in any dataset will almost always be measured in various units, such as double meaning words in the messages.

**Centroid-based Clustering**

-The algorithm will then repeat the process until the centroids do not change. [centroid˙based˙clustering} To repeat k means multiple times with different initializations and select the best result, but for instance, Centroid is difficult to interpret if we are dealing with text type of data because it can’t be interpretable. A better approach to this type of situation would be to use the K Medoids algorithm.

Centroid-based clustering seeks to divide a dataset into discrete groups or clusters. Centroids, or the central points within each cluster, are the concepts that define these clusters. In essence, centroid-based clustering is an iterative clustering technique in which clusters are created depending on how closely data points are clustered around the centroid.

For Ex- [K – means algorithm](https://www.geeksforgeeks.org/k-means-clustering-introduction/) is one of the popular examples of this algorithm.

