**Dataset description:**

The dataset contains 14 attributes published by the Cleveland database. The data set refers to heart disease in patients. Where takes patient info and valued from 0 that means no presence to 4.

The dataset attributes are:

* age
* sex
* chest pain type (4 values)
* resting blood pressure
* serum cholestoral in mg/dl
* fasting blood sugar > 120 mg/dl
* resting electrocardiographic results (values 0,1,2)
* maximum heart rate achieved
* exercise induced angina
* oldpeak = ST depression induced by exercise relative to rest
* the slope of the peak exercise ST segment
* number of major vessels (0-3) colored by flourosopy
* thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

**Data preprocessing:**

The data has no empty or missing values. It was already preprocessed.

**K-Means Algorithm:**

K-Means is used to divide data objects into clusters it works by measuring the distance between each point and group them together in one cluster the distance is measured by Manhattan distance.

**Hieratical Algorithm:**

is an algorithm that adds data in hierarchical cluster model. By assigning data points to cluster of their own. And then merge the closest clusters into the same cluster. Which at the end only one cluster remains

**Comparison:** Hieratical clustering cannot handle bigdata as the time complexity is quadratic O(n^2), but the K-Means can handle big data as the time complexity is linear O(n). K-Means uses steps to be able to visualize data by doing number of steps as following:

1. specifying number of clusters
2. randomly assign each data point to cluster
3. compute the cluster centroids.
4. Reassigning each point to the closest cluster
5. Re-computer cluster centroids
6. Then repeating steps 4 and 5 until there is no improvements into clusters

At the Hieratical clustering can be shown as dendrogram. We start at the bottom with some certain number of points each one of them assigned to different cluster. The closest ones are merged till at the end we have only one cluster at the top.

The K-Means cluster accuracy alone can be 78% but it can be enhanced by better model to give accuracy of 95%.