

# Data Driven Computing and Networking (DDCN-2019)

## Centroid Clustering using k-Means Algorithm

### **A. Write a python script to perform the following tasks:-**

1. Load all the required packages to implement k-Means Clustering algorithm
2. Create two lists named x and y having values {1,5,1.5,8,1,9} and {2,8,1.8,8,0.6,11} respectively.
3. Display scatter chart of x and y
4. Create an array X which stores the pair of each element of x and y respectively.
5. Apply KMeans function with two number of clusters and store its output in variable kmeans, representing a clustering model.
6. Fit kmeans clustering model on array X.
7. Extract centroids and labels from the model kmeans and print them on console.
8. Define color list having two different colors red and green to represent two clusters
9. For each element of the array X
  - a. Print coordinates and labels along with the element of X
  - b. Plot each element of X using colors and labels
10. Plot centroids of both clusters
11. Display scatter chart showing all elements of X with designed clusters in specified colors.
12. Import “pandas” package” and copy dataset “faithful.csv” to the Destop folder on your system
13. Open dataset file “faithful.csv” and store it a variable “d”
14. Print first five records of the variable “d”
15. Plot scatter chart of columns “eruptions” and “waiting” of the variable “d”
16. Show scatter chart with chart title as “Old Faithfull Data Scatter Plot”, x axis as “Length of eruptions” , and y axis as “Time between eruptions” .
17. Create an array “d1” which stores the elements of the variable “d”.
18. Apply KMeans function with two number of clusters and store its output in variable kmeans, representing a clustering model.
19. Fit kmeans clustering model on array “d1”.
20. Extract centroids and labels from the model kmeans.
21. For each element of the array “d1”
  - a. Extract observations for each level from the array “d1” and store it in variable “ds”
  - b. Plot both columns from the variable ds and centroids for each cluster
  - c. Increase the size of centroid points
22. Display scatter chart showing all elements of the datasets with designated clusters and centroids.

## **Hierarchical Clustering using Mean Shift Algorithm**

### **B. Write a python script to perform the following tasks:-**

1. Load all the required packages to implement Hierarchical Clustering using MeanShift algorithm
2. Import packages to generate sample data
3. Define center points
4. Generate sample of data sets and store it in X,Y
5. Display scatter chart of generated sample data (X)
6. Apply MeanShift function and store its output in variable kmeans, representing a clustering model.
7. Fit generated clustering model on the data X.
8. Extract centroids and labels from the cluster model.
9. Extract number of clusters from the cluster model and print the number of clusters on console.
10. Define list of colors
11. Print colors and labels
12. For each observation of data X
  - a. Plot observations with suitable colors according to the designated labels
13. Plot scatter chart of cluster centers and mark them with “x”
14. Display scatter chart