

## Assignment 9 : Clustering

Download the following dataset :

<https://www.kaggle.com/datasets/erdemtaha/cancer-data/data>

- 1 Drop Id and Diagnosis columns
- 2 Cluster the data using K means clustering :
  - a. Determine optimal number of clusters (between 2-10) using
    - i. Elbow method with Inertia and
    - ii. Silhouette Analysis
  - b. Use the diagnosis column to calculate homogeneity score and evaluate quality of clustering
- 3 Cluster the data using DB scan clustering.
  - a. Determine appropriate values of eps and min\_samples parameters
  - b. Use the diagnosis column to calculate homogeneity score and evaluate quality of clustering

Submission Guideline:

Submit your assignment by EoD Wednesday 16<sup>th</sup> April .