Assignment 2

Machine Learning (**CS564**)

Date:- 06-092023 **Deadline:-13-09-2023**

Instructions:

1. **Coding must be done using python and you are not supposed to use already available libraries of clustering algorithms.**
2. All the assignments should be completed and uploaded before the deadline.
3. Markings will be based on the correctness and soundness of the outputs. Marks will be deducted in case of plagiarism.
4. Proper indentation and appropriate comments are mandatory.
5. You should zip all the required files and name the zip file as ***roll_no*.zip**, eg. **1501cs11.zip. 6.** Upload your assignment (**the zip file**) in the following link: https://docs.google.com/forms/d/e/1FAIpQLSc_KZt3Cgnsno1xZRrwa2awF_T8JXaFk5A2qi2EICsPItihvA/viewform?usp=sf_link

Assignment: Cluster the provided dataset "cancer.csv" of cancer patients using hierarchical agglomerative and DB-scan clustering algorithms. Plot the clusters and count the number of points belonging to each cluster. Each cluster must be plotted using different colors. The clustering has to be done using all the attributes except "id" and "diagnosis" but for plotting purposes, you are supposed to use "radius_mean on X-axis" and "texture_mean on Y-axis".

For iii) **hierarchical agglomerative** : Plot three different clusters for the following parameters and observe the difference. Number of clusters = 2,
Linkage= single
Linkage= complete
Linkage= average

For iv) **DB-scan** : Plot three different clusters for the following parameters and observe the difference.
Eps = 0.2, minPoints = 6

Eps = 0.5, minPonts = 6
Eps = 0.2, minPoints = 3