LAB (4%), Due on 22nd October at 3 pm

You are provided with a medical dataset (two files Clusttering_Data, Classification_Data) with a total 17 features. You are required to demonstrate your ability to create a Semi-Supervised Learning solution based on the dataset.

The two dataset are to be used as following:

- Clustering_Data: Use this for question a below. I.e. use this to create clusters.
- Classification_Data: Assing labels to this dataset after training a kNN classification models using the clusters from ***question a***.

Task

Generate a Jupyter Notebook that clearly demonistreate your approach to:

a. Clustering the dataset into 3 or 4 or 5 clusters using scikit-learn **k-Means** . After clustering, create a new column on the dataframe and assing clusters for each row in the dataframe. This column will then become your target colum for the next question

b. Using your final daframe from question a, **train a kNN classification model (k = 7).** Also, s**et the the distance methods used to be Euclidean**. Use this model to assing labels to the Classifcation_Data.

Additional

- Submit one Jupyter Notebook file (solution for the tasks mentioned above). Submitting a file of a different type will lead to a zero score.
- Name your file with your admission number and group, e.g. **A-111111.ipynb** or **B-111111.ipynb**
- **Submit your file using the correct submission link on E-Learning.** *If by some random chance E-Learning fails, the various class reps. can reach out. I'll then provide an alternative Dropbox Link or Google Drive Link to ensure everyone can submit their work successfully.*
- **Do not send any files via email** as these may be labeled as spam.
- Take note of the liberty to pick a K value for clusttering.