

Tools and Software Data Science  
LAB Assignment no. 1  
Group Work  
Due Date: 31/10/2023  
Dr. Mohamed Abdelhafeez

## Objectives

This assignment aims to make students familiar with python programming and to begin the exploration process of a dataset.

## Download and Read Data

Each group should deal with two datasets:

1. text data such as Subset of reuters dataset at <https://paperswithcode.com/dataset/reuters-21578>
2. Tabular data such as IRIS at <https://archive.ics.uci.edu/ml/datasets/iris>

## Visualization

## Notes

Marks are put on effort and trials so copied assignments will be penalized, so not delivering the assignment would be much better.

## Tasks

For each dataset:

1. You should deliver a report which contains:
  - Explanation of the dataset attributes and details in a tabular form.
  - The visualization and with comments on them if available.
2. You should write your code in python and you may use scikit-learn, numpy, scipy or pyplot whenever needed.
3. It is preferred to use [Google Colab](#) in this sheet and any coding assignment.
4. Any further effort (more visualizations, analysis, etc.) will be appreciated.
5. Create csv file that contain the dissimilarity matrix