AERO 4 - MATHEMATICAL TOOLS FOR DATA SCIENCE (2023/2024)

Instructions:

Provide:

- A GitHub repository.
- A file in .pdf format presenting the work carried out. This involves explaining the problem, the possible solutions and those you have chosen by presenting the strengths and weaknesses. Particular attention should be paid to the explanation of the considered algorithms.
- A source file containing the properly commented Python code.

Please note: Projects that do not comply with one of these rules will receive a 20% penalty.

You can send **one individual solution** to: Leila.gharsalli@ipsa.fr by January the 10 at the latest.

Subject:

You have a database *'data.npy'* that contains 3879 examples with 18 features each. Each example corresponds to an aircraft trajectory with its position in the sky and other significant features. The objective here is to study unsupervised clustering techniques that allow to group the examples in the data set. You should define a metric to decide if your method is performing good. In addition, the number of clusters in the data set needs to be determined. Any method suggestion not seen mainly in the course will be rewarded.

The data are clean, and you don't need to pre-process them. But if not, what method would you suggest for data processing ? Justify your answer.

Good luck!