

## SECOND SESSION

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

#### Instructions:

You should 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](mailto:Leila.gharsalli@ipsa.fr) by June the 17<sup>th</sup> at the latest.

#### Subject:

You have a database, already splitted into train and test data, which contains a survey on air passenger satisfaction. In fact, following the pandemic, the airline industry suffered from a massive setback. Hence, in order to revitalize the industry in the face of the current recession, it is necessary to understand the customer pain points and improve their satisfaction with the services provided. You are facing a prediction problem where it is necessary to determine which of the following levels of satisfaction with the airline the passenger belongs to: Satisfaction, Neutral or dissatisfied. Start by analyzing the given train and test datasets then propose an adequate Machine Learning model capable of classifying what factors are highly correlated to a satisfied (or dissatisfied) passenger and predicting the latter' satisfaction. Any method suggestion not seen mainly in the course will be rewarded.

Good luck!