ST451: Assessed Coursework - Project

You will undertake a project that will determine your final mark of the course by 50 per cent. The project will require you to analyse one or more real world datasets of your choice. You can use the OpenML website, the UCI repository, or any other open access domain.

The project will consist of two tasks than can be chosen from the following

1. Regression: where the problem consists of continuous target variable(s).
2. Classification: where problem consists of categorical target variable(s).
3. Cluster Analysis: where the problem consists of identifying homogeneous population groups that can then be used in the context of the empirical application.
4. Sequential data: Any of the above settings where the data become available in a sequential manner and need to be incorporated in the decision-making process.

You will be expected to present the empirical problem, consider and implement competing methods to use the available data to address it. At least one of the adopted methods should be among the topics covered in the course. It is also fine if more than one or even all of the adopted methods are among the topics covered, as long as they are relevant for the empirical problem.

In all cases, the output from these techniques should be described in non-technical language targeting people with a minimal quantitative background.

The results of the project should be presented in a 10-page article in A4 format. The 10-page limit includes figures and tables but excludes the title page, table of contents and references. In addition to the 10-page article, which should be submitted via a soft copy, your Python code should also be submitted with appropriate comments and description via a Jupyter notebook.

The project is due **Wednesday, May 8th noon.**